

# UC San Diego

## UC San Diego Previously Published Works

### Title

Bacterial and Archaeal Viruses of Himalayan Hot Springs at Manikaran Modulate Host Genomes

### Permalink

<https://escholarship.org/uc/item/0wv5q8js>

### Authors

Sharma, Anukriti  
Schmidt, Matthias  
Kiesel, Bärbel  
et al.

### Publication Date

2018

### DOI

10.3389/fmicb.2018.03095

Peer reviewed



# Bacterial and Archaeal Viruses of Himalayan Hot Springs at Manikaran Modulate Host Genomes

Anukriti Sharma<sup>1,2,3</sup>, Matthias Schmidt<sup>4</sup>, Bärbel Kiesel<sup>5</sup>, Nitish K. Mahato<sup>1</sup>, Lauren Cralle<sup>2,3</sup>, Yogendra Singh<sup>1</sup>, Hans H. Richnow<sup>4</sup>, Jack A. Gilbert<sup>2,3</sup>, Wyatt Arnold<sup>3</sup> and Rup Lal<sup>1\*</sup>

<sup>1</sup> Department of Zoology, University of Delhi, New Delhi, India, <sup>2</sup> Biosciences Division, Argonne National Laboratory, Lemont, IL, United States, <sup>3</sup> Department of Surgery, University of Chicago, Chicago, IL, United States, <sup>4</sup> Department of Isotope Biogeochemistry, Helmholtz Centre for Environmental Research, Leipzig, Germany, <sup>5</sup> Department of Environmental Microbiology, Helmholtz Centre for Environmental Research, Leipzig, Germany

## OPEN ACCESS

### Edited by:

Eric Altermann,  
AgResearch (New Zealand),  
New Zealand

### Reviewed by:

Joao Carlos Setubal,  
University of São Paulo, Brazil  
William C. Nelson,  
Pacific Northwest National Laboratory  
(DOE), United States

### \*Correspondence:

Rup Lal  
ruplal@gmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 21 August 2018

**Accepted:** 29 November 2018

**Published:** 14 December 2018

### Citation:

Sharma A, Schmidt M, Kiesel B,  
Mahato NK, Cralle L, Singh Y,  
Richnow HH, Gilbert JA, Arnold W  
and Lal R (2018) Bacterial  
and Archaeal Viruses of Himalayan  
Hot Springs at Manikaran Modulate  
Host Genomes.  
*Front. Microbiol.* 9:3095.  
doi: 10.3389/fmicb.2018.03095

Hot spring-associated viruses, particularly the archaeal viruses, remain under-examined compared to bacteriophages. Previous metagenomic studies of the Manikaran hot springs in India suggested an abundance of viral DNA, which prompted us to examine the virus–host (bacterial and archaeal) interactions in sediment and microbial mat samples collected from the thermal discharges. Here, we characterize the viruses (both bacterial and archaeal) from this Himalayan hot spring using both metagenomics assembly and electron microscopy. We utilized four shotgun samples from sediment (78–98°C) and two from microbial mats (50°C) to reconstruct 65 bacteriophage genomes (24–200 kb). We also identified 59 archaeal viruses that were notably abundant across the sediment samples. Whole-genome analyses of the reconstructed bacteriophage genomes revealed greater genomic conservation in sediments (65%) compared to microbial mats (49%). However, a minimal phage genome was still maintained across both sediment and microbial mats suggesting a common origin. To complement the metagenomic data, scanning-electron and helium-ion microscopy were used to reveal diverse morphotypes of Caudovirales and archaeal viruses. The genome level annotations provide further evidence for gene-level exchange between virus and host in these hot springs, and augments our knowledgebase for bacteriophages, archaeal viruses and Clustered Regularly Interspaced Short Palindromic Repeat cassettes, which provide a critical resource for studying viromes in extreme natural environments.

**Keywords:** archaeal viruses, bacteriophages, scanning-electron microscopy, helium-ion microscopy, Caudovirales

## INTRODUCTION

Viruses are the most prominent predators and mediators of genetic transmission in prokaryotic communities in extreme thermal environments, which are characterized by lack of eukaryotes (Breitbart et al., 2004). The majority of well-characterized prokaryotic viruses belong to the order Caudovirales, which are bacteriophages with tails (Krupovic et al., 2011). In the past, the sequence-based viral characterization as well as electron and ion microscopy have been used for elucidating viral diversity within geothermally heated ecosystems (Häring et al., 2005).

A series of Himalayan hot springs (surface temperature > 95°C) located at an altitude of 1,760 m at Manikaran (32°02'N, 72°21'E) have previously been investigated using sequencing approaches to characterize the bacterial communities (Dwivedi et al., 2012; Mahato et al., 2014; Sharma et al., 2014, 2016b; Tripathi et al., 2016). Culture-independent analysis of the sediment and microbial mat samples demonstrated unexpectedly high microbial diversity in these hot springs (Sangwan et al., 2015; Sharma et al., 2016a). Sediment samples with temperatures ranging from 78 to 98°C were dominated by archaeal genotypes, while microbial mats maintained high abundance of resident bacterial populations with integrated phage DNA (Sangwan et al., 2015). However, the role of these viruses in microbial community dynamics remains undetermined.

We explored the virome of these Himalayan hot springs by coupling metagenomic profiling of the bacteriophages and archaeal viruses with *in situ* electron microscopy of virus-enriched sediment and microbial mat samples. We assembled four metagenomes from sediments at a range of temperatures [98°C (MnS1), 85°C (MnS2), 81°C (MnS3), 85°C (MnS4)], and two from microbial mat samples (MM1, MM2; 50°C), in order to reconstruct 65 phage genomes. Moreover, 59 different archaeal viruses were found to be abundant in the sediment samples. We also reconstructed Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) cassettes, and the genomes of potential bacterial hosts. This allowed us to (1) determine and characterize the potential phage ensemble at the Manikaran hot springs, and (2) explore the gene contribution of phages to their bacterial and archaeal hosts. Furthermore, the use of electron microscopy on sediment and mat samples allowed us to visualize and make cladistics assignments to the prokaryotic virosphere especially the archaeal viruses since the archaeal reference sequence databases still remain limited (Håring et al., 2005).

## RESULTS AND DISCUSSION

### Viral Diversity in Microbial Mat and Sediment Samples

The Himalayan hot springs located at Manikaran were analyzed for prokaryotic viral genomic signatures. These hot springs are characterized by abundance of Fe (12,380 ug g<sup>-1</sup>), Mn (2,189 ug g<sup>-1</sup>), As (80 ug g<sup>-1</sup>) and Se (1.5 ug g<sup>-1</sup>). Previously we demonstrated that the microbial mats at Manikaran are enriched in bacterial genera such as *Enterobacter*, *Bdellovibrio*, *Clostridium*, and *Achromobacter* (Sangwan et al., 2015). Furthermore, we reconstructed the genome of a novel *Enterobacter* strain with its corresponding bacteriophage (Sangwan et al., 2015), suggesting a potential phage–host dynamics at the Manikaran hot springs, which became the focus of the current study.

Metagenomic data from the microbial mat ( $n = 2$ ; 50°C) and sediment samples ( $n = 4$ , 78–98°C) (Table 1 and Supplementary Figure S1) was assembled for binning phage genomes. We expected differential viral enrichments

in the moderately mesophilic microbial mats compared to the relatively thermophilic sediment samples (78–98°C), which were dominated by Archaea (Sangwan et al., 2015). While water samples (95°C) were collected, we were unable to detect sufficient phage DNA (based on BLAST results against viral database) in the metagenomic sequence data, and thus water samples were excluded from the analyses.

In total, 709 contigs were identified as having sequence similarity to bacteriophage, 103 of which were found in microbial mat samples, while 606 originated in the sediment samples (Table 1). The contigs were then re-assembled and curated into 65 bacteriophage genomes (Table 2), of which 90% were assigned to the order Caudovirales. Caudovirales are the most abundant and well characterized viruses (Krupovic et al., 2011), as a result, extensive reference databases exist for the analysis and annotation (Munson-McGee et al., 2018). The majority (71%; 10/14) of bacteriophage genomes from microbial mat samples belonged to the family *Myoviridae*, whereas genomes reconstructed from the sediment samples were assigned either to *Myoviridae* (39%; 20/51), *Siphoviridae* (35%; 18/51), or *Podoviridae* (16%; 8/51) (Figure 1A and Table 2). *Siphoviridae* were most abundant in the high temperatures (sediments), while *Myoviridae* phages were most populous at samples with lower temperatures (microbial mats) (Table 2).

Using a database of 95 archaeal virus genomes from NCBI, we were able to identify metagenomic contigs associated with 59 different archaeal viruses, which originated primarily from the sediment samples (Table 3 and Supplementary Figure S2), corresponding to the dominance of Archaea in this environment. At the family level, these contigs were assigned to the archaeal virus families of *Fuselloviridae*, *Siphoviridae*, and *Lipothrixviridae* (Table 3 and Supplementary Figure S2). The detailed annotations of these archaeal virus contigs are mentioned in Table 3. This is only the second study to report a relatively novel archaeal turreted icosahedral virus with *Metallosphaera* (MT1V1 strain) as the archaeal host, which was recently isolated from Yellowstone National Park (YNP) (Wagner et al., 2017) (Table 3). While, we were able to reconstruct draft genomes for bacteriophages, we were not able to successfully assemble draft archaeal virus genomes potentially due to low sequencing depth and lack of biological replicates (Albertsen et al., 2013; Nielsen et al., 2014; Sangwan et al., 2016).

Electron microscopy and helium-ion microscopy (HIM) was used to visualize viruses in these samples. HIM revealed virus-like particles of 50–90 nm attached to the surface of bacteria (Figure 1B). HIM was able to capture good resolution images even though we used heterogeneous mat and sediment samples directly. These images paralleled what was seen in electron micrographs obtained from virus–host infections in pure-cultures (Rice et al., 2001; Rachel et al., 2002; Håring et al., 2005). These results led us to further explore pure virus enrichments from sediment and microbial mat samples, which could then be examined even under the scanning electron microscope (SEM) at a high enough resolution to refine the taxonomic assignments.

The spindle-shaped viruses were found in the sediment samples, which were characterized by diverse range of sizes. These were then assigned to the family *Fuselloviridae* based

**TABLE 1** | Characteristics of the metagenome samples analyzed from the Manikaran hot springs.

S. No	Site	Assembly size (Mb)	Temp. (°C)	pH	Bacteriophages reconstructed	Size of the largest bacteriophage (kb)	CRISPR cassettes	Number of archaeal viral contigs
1	MM1	95	50	7	5	200	45	70
2	MM2	13	50	7	9	104.9	7	52
3	MnS1	64	98	7.3	28	42.2	307	2713
4	MnS2	53	85	7.1	6	24.3	54	1400
5	MnS3	18	81	7.1	7	37.5	17	1180
6	MnS4	74	78	7.6	11	25.3	77	5132

on their morphology (**Figures 1C,D**) (Prangishvili, 2013). This family was found to be abundant across sediment samples based on metagenomic DNA as well. However, we were not able to visualize other families such as *Siphoviridae* and *Lipthrixviridae* (identified using sequence-based analyses) suggesting that our current culturing methodology was not able to enrich these viral families. The smallest archaeal viral particle observed via electron microscopy was 17 nm in width and 60 nm in length, while the largest was 40 nm in width and 200 nm in length (data not shown). Interestingly, a high resolution magnified image of one of these viral particles revealed two different asymmetric ends, which is consistent with reports indicating that archaeal viruses belonging to family *Fuselloviridae* have a ‘mouth’ and ‘tail’ (host attachment and penetration is achieved through the ‘mouth’) (**Figure 1D**) (Prangishvili et al., 2013). The electron microscopy of the four sediment samples revealed an abundance of archaeal viruses of a similar spindle-shaped morphotype. We further visualized all sediment enrichments and the same spindle shaped *Fuselloviridae* morphotype was always present (two examples are shown in **Figure 1D**).

The microbial mat samples were characterized by a wide variety of bacteriophages belonging to different tailed (*Myoviridae*, *Siphoviridae*, *Podoviridae*) and tailless (single and double membrane bound viral particles) families (**Figure 1E**) (Clokic et al., 2011). These different families each have their own characteristic morphology, with *Myoviridae* particles having a head size of 50–110 nm (**Figure 1E**), *Siphoviridae* phages having a head size of 50–70 nm (**Figure 1E**), and *Podoviridae* phages possessing a very small tail that is difficult to observe via scanning electron microscopy (Clokic et al., 2011) (though we were still able to image a *Podoviridae* particle) (**Figure 1E**, element 5). This diversity of bacteriophages within microbial mats is perhaps best attributed to the moderately mesophilic (up to ~50°C) environments, that allow for more bacterial host diversity (Pietilä et al., 2014). Archaea are better at surviving the extreme temperatures found in the sediment samples, and so outcompete bacteria (Munson-McGee et al., 2018), resulting in an increased abundance of archaeal viruses.

## Phylogenomic Clustering Analysis

As phage genomes lack a credible set of marker genes, we relied on the total predicted proteome (based on the complete genome sequence data of each phage) for phylogenomic delineation

(Rohwer and Edwards, 2002). Sediment samples showed overall site-specific clustering, whereas microbial mat phages were distributed throughout the phylogeny (**Supplementary Figure S3**), which is supported by the greater genomic conservancy in the sediment samples. Among the sediment samples, phage genomes associated with the MnS1 clustered together. The phage genomes recovered from microbial mats were not seen to cluster together in the proteome-based phylogenetic tree, presumably due to the relatively low number of phage (nodes) representatives from mats in comparison to the sediment samples. Interestingly, the family level assignments for both microbial mat and sediment phage genomes were not tight clades, but rather were distributed throughout the entirety of the phylogenetic tree (**Supplementary Figure S3**). However, small sub-clades were observed for the family *Myoviridae* (**Supplementary Figure S3**). This could potentially be due to a greater diversity of phage genomes, as well as a result of the lack of a sound methodology for reconstructing phage phylogenies (Hatfull et al., 2010). Additionally, this could be attributed to incomplete databases, as only a small number of phage genomes have complete family level characterization, or to the potential novelty of the phage data analyzed in this study.

## Functional Characterization of Phage Genomes

Functional annotation was used to predict the lifestyles of the assembled bacteriophage genomes using the Phage Classification Tool Set (PHACTS) (McNair et al., 2012). The proteins involved in integration, excision, lysogeny, regulation of expression and toxins genes are predominantly important toward classifying temperate phages, whereas proteins involved in nucleotide metabolism, phage lysis and structural Proteins are predominantly important toward classifying virulent phages (McNair et al., 2012). Using supervised learning based Random Forest classifier PHACTS can predict the phage genome lifestyle with an average accuracy of 99%. The reconstructed MM1 phages were predicted to have temperate (44%) and lytic (56%) lifestyles (**Table 2**). However, all the reconstructed MM2 phages were predicted to be temperate in lifestyle (100%). Meanwhile, sediment phages were predicted to be both lytic (MnS1 = 50%, MnS2 = 67%, MnS3 = 57%, and MnS4 = 63%) and temperate (MnS1 = 50%, MnS2 = 33%, MnS3 = 43%, and MnS4 = 37%) (**Table 2**). These results are, however, based

**TABLE 2** | Characteristics of phages reconstructed across microbial mat and sediment of Manikaran hot springs.

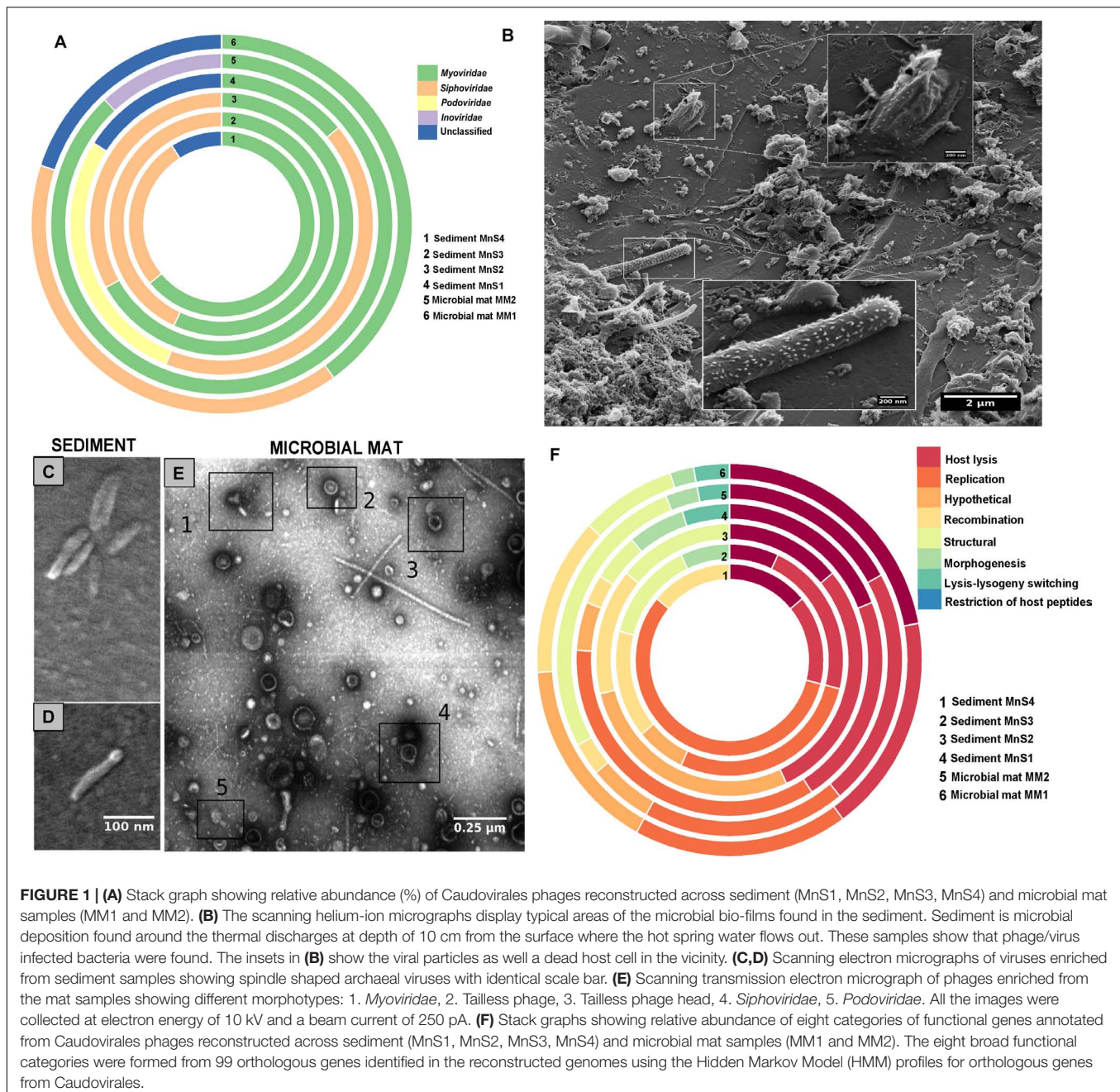
Microbial Mat-MM1				Microbial Mat-MM2							
Phage size (in kb)	No. of contigs	Average coverage, %GC	Host taxa	Phage family	Phage lifestyle	Phage size (in kb)	No. of contigs	Average coverage, %GC	Host taxa	Phage family	Phage lifestyle
200	4	11, 60.4	Unknown	U	T	42.6	1	227, 64.5	(g) <i>Ralstonia</i>	M	T
49	2	20, 54.1	(o) Enterobacteria	S	T	14.4	3	340, 57.3	(g) <i>Ralstonia</i>	I	L
42	6	24, 55.3	(g) <i>Enterobacter</i>	M	T	30.3	4	10, 65.2	(g) <i>Ralstonia</i>	M	L
48	5	5, 48.4	(g) <i>Enterobacter</i>	M	T	12.5	7	220, 57.5	(g) <i>Ralstonia</i>	M	L
38	8	13, 49.6	(g) <i>Caulobacter</i>	S	T	16.9	3	11, 62.3	(g) <i>Burkholderia</i>	M	T
<b>Sediment – MnS1</b>						13.2	3	6, 63.4	(g) <i>Burkholderia</i>	M	L
20.5	5	50, 60.6	(g) <i>Burkholderia</i>	M	T	104.9	44	212, 62.6	(g) <i>Ralstonia</i>	M	L
19.23	1	16, 43.4	(g) <i>Streptomyces</i> , <i>Rhodococcus</i> , <i>Corynebacterium</i>	S	L	41	2	12, 62.5	(g) <i>Burkholderia</i>	M	T
42.22	1	16, 62.7	(g) <i>Bordetella</i>	P	L	12.5	11	5, 61.3	(g) <i>Herbaspirillum</i>	M	T
17.2	1	10, 63.3	(g) <i>Geobacillus</i>	U	T	<b>Sediment – MnS2</b>					
12.6	2	240, 57.1	(g) <i>Rhizobium</i>	S	T	24.3	33	3, 66.8	(o) Enterobacteria	M	L
18.9	13	13, 60.1	(g) <i>Stenotrophomonas</i>	S	L	11.4	10	8, 71	(g) <i>Burkholderia</i>	M	L
20.3	3	14, 58.8	(g) <i>Pseudomonas</i>	P	L	11.8	16	4, 68.9	(g) <i>Bacillus</i>	M	L
20.1	6	18, 63.1	(g) <i>Rhizobium</i>	S	T	10.4	12	4, 71.3	(g) <i>Bacillus</i>	M	T
15	1	34, 58.1	(o) Enterobacteria	S	T	10.1	14	4, 66.5	(g) <i>Pseudomonas</i>	S	T
12.5	1	8, 63.9	(g) <i>Synechococcus</i>	U	T	13.7	19	3, 69.1	(g) <i>Synechococcus</i>	S	L
10.3	11	4, 51.2	(g) <i>Pseudomonas</i>	P	T	<b>Sediment – MnS3</b>					
15	1	8, 69.5	(g) <i>Ruegeria</i>	S	T	37.5	4	31, 63.8	(g) <i>Pseudomonas</i>	S	T
11.4	15	5, 59.1	(g) <i>Burkholderia</i>	P	L	10.1	14	4, 64.7	(g) <i>Rhodococcus</i>	M	L
16	8	42, 59.2	(g) <i>Burkholderia</i>	P	L	10.6	11	3, 67.1	(g) <i>Bacillus</i>	S	L
14.5	11	11, 58.1	(g) <i>Synechococcus</i>	S	L	10.1	17	3, 65.8	(g) <i>Prochlorococcus</i>	M	T
11.3	7	7, 59.1	(g) <i>Mycobacterium</i>	S	T	12.7	25	3, 65.1	(g) <i>Roseobacter</i>	S	L
10.4	6	22, 60.2	(g) <i>Pseudomonas</i>	S	T	10	14	3, 62.9	(g) <i>Klebsiella</i>	M	T
24.8	20	13, 64.1	(g) <i>Synechococcus</i>	S	L	36.8	20	4, 62.8	(g) <i>Prochlorococcus</i>	M	L
10.01	14	3, 56.7	(o) Enterobacteria	M	L	<b>Sediment – MnS4</b>					
10.4	13	5, 58.5	(g) <i>Bacillus</i>	M	L	25.3	9	7, 66.8	(g) <i>Synechococcus</i>	M	L
27.9	25	12, 61.4	(g) <i>Salmonella</i>	P	L	12.6	2	8, 62.8	(g) <i>Mycobacterium</i>	S	L
23.02	16	17, 59.9	Unknown	U	L	10.9	3	6, 64.8	(g) <i>Synechococcus</i>	M	L
15.9	26	5, 66.7	(g) <i>Yersinia</i>	M	L	20.1	7	4, 67.7	(g) <i>Mycobacterium</i>	S	T
12.7	1	6, 57.8	(g) <i>Brucella</i>	P	L	18.3	12	10, 68.8	(g) <i>Sinorhizobium</i>	M	T
35.6	1	9, 59.7	(g) <i>Burkholderia</i>	P	T	23.1	22	5, 62.9	(g) <i>Roseobacter</i>	S	T
13.6	5	7, 67.1	(g) <i>Paracoccus</i>	S	T	13.7	5	3, 65.8	(g) <i>Thermus</i>	U	L
21.07	26	3, 62.4	(g) <i>Streptomyces</i>	S	T	21.9	31	4, 69.8	(g) <i>Cronobacter</i>	M	L
						12	23	3, 65.7	(g) <i>Synechococcus</i>	M	L
						20.7	21	3, 68.2	(o) Enterobacteria	M	T
						20.7	21	3, 68.3	(o) Enterobacteria	M	L

Phage family is denoted by symbols M – Myoviridae, S – Siphoviridae, I – Inoviridae, P – Podoviridae, U – not classified. Phage lifestyle is denoted by symbols T – Temperate and L – Lytic. The host taxa assigned at genus level begin with (g) and ones assigned at order level begin with (o).

on the reconstructed phage genomes only and thus cannot represent the lifestyle of total bacteriophage community in these samples.

Hidden Markov Model (HMM) profiles for orthologous genes from *Caudovirales* were used to assign function to the genes in the assembled genotypes. In total, 99 orthologous genes were

annotated and further clustered into eight functional categories: host lysis, replication, hypothetical proteins, recombination, structural, morphogenesis, lysis-lysogeny switching, and restriction of host peptides (**Supplementary Table S1** and **Figure 1F**). All the phage genomes presented hypothetical proteins (27% ± 12), as well as genes involved with replication



(21% ± 4), host lysis (15% ± 4), morphogenesis (12% ± 8), recombination (11% ± 9), and structure (10% ± 5) (**Figure 1F**).

## Comparative Genomics of Reconstructed Phage Genomes

The phage genomes ( $n = 65$ ) were compared based on whole genome alignments using dot plots and network plots based on tetranucleotide frequencies. Low sequence similarity was observed between the phages assembled from the microbial mats ( $49 \pm 1.6\%$ ) and high sequence similarity ( $65 \pm 1\%$ ) was observed for those assembled from the sediment (**Figures 2A, B**).

Additionally, we identified differentially abundant ( $p < 0.05$ , Welch's  $t$ -test) functional gene orthologs between phage genomes from microbial mats and sediment samples (**Figure 2C**). The structural phage proteins including packaging chaperon FI (gpFI) and FII (gpFII) proteins, phage late control gene D protein, and CI repressor protein (**Figure 2C**) were higher ( $p < 0.05$ ) in abundance in microbial mats. CI repressor proteins play significant role in phages by determining the lifestyle, i.e., lytic or lysogenic (Pedersen et al., 2010).

A set of shared orthologs were found between sediment ( $n = 4$ ) and microbial mat ( $n = 2$ ) phage genomes, including genes coding for structural phage proteins, such as phage

**TABLE 3** | The BLASTX results of metagenome contigs from sediments and mat samples, searched against the archaeal virus sequence database.

S. No	Best blast hit (BBH) annotation of contigs	Archaeal virus family	MNS1 (# of BBH, average %identity)	MNS2 (# of BBH, average %identity)	MNS3 (# of BBH, average %identity)	MNS4 (# of BBH, average %identity)	Mat (# of BBH, average %identity)
1	<i>Halovirus</i> HHTV-1	<i>Fuselloviridae</i>	31, 90	25, 89	13, 91	53, 88	7, 90
2	<i>Acidianus</i> bottle-shaped virus 2 strain ABV2	<i>Lipothrixviridae</i>	73, 55	31, 60	18, 51	110, 57	3, 61
3	<i>Acidianus</i> bottle-shaped virus 3 strain ABV3	<i>Lipothrixviridae</i>	92, 57	30, 54	26, 60	165, 56	3, 57
4	<i>Acidianus</i> filamentous virus 1	<i>Lipothrixviridae</i>	51, 60	11, 61	27, 59	71, 58	1, 60
5	<i>Acidianus</i> filamentous virus 2	<i>Lipothrixviridae</i>	78, 52	37, 54	28, 58	151, 55	2, 51
6	<i>Acidianus</i> filamentous virus 3, partial viral genome	<i>Lipothrixviridae</i>	0	24, 60	12, 56	119, 61	2, 59
7	<i>Acidianus</i> filamentous virus 6	<i>Lipothrixviridae</i>	72, 62	34, 61	13, 63	112, 64	4, 61
8	<i>Acidianus</i> filamentous virus 7	<i>Lipothrixviridae</i>	68, 70	31, 72	13, 71	119, 70	9, 71
9	<i>Acidianus</i> rod-shaped virus 1	<i>Lipothrixviridae</i>	53, 52	23, 51	14, 50	97, 53	7, 54
10	<i>Acidianus</i> rod-shaped virus 2 strain ARV2	<i>Lipothrixviridae</i>	94, 51	53, 56	48, 49	154, 42	7, 56
11	<i>Acidianus</i> spindle-shaped virus 1	<i>Lipothrixviridae</i>	46, 59	20, 53	12, 51	92, 67	3, 53
12	<i>Acidianus</i> two-tailed virus	<i>Lipothrixviridae</i>	94, 51	47, 62	70, 63	145, 53	4, 61
13	<i>Aeropyrum permix ovoid</i> virus 1 complete genome	<i>Clavaviridae</i>	0	3, 58	0	17, 62	4, 67
14	<i>Aeropyrum permix</i> spindle-shaped virus 1 complete genome	<i>Clavaviridae</i>	0	12, 55	0	43, 58	1, 49
15	Archaeal BJ1 virus	<i>Siphoviridae</i>	23, 58	5, 61	3, 67	35, 68	3, 63
16	<i>Haloarcula hispanica</i> icosahedral virus 2	<i>Fuselloviridae</i>	17, 59	8, 63	10, 62	21, 61	6, 55
17	<i>Haloarcula hispanica</i> pleomorphic virus 1	<i>Fuselloviridae</i>	3, 67	3, 55	5, 63	8, 61	8, 50
18	<i>Haloarcula hispanica</i> pleomorphic virus 2	<i>Fuselloviridae</i>	3, 59	1, 62	2, 57	15, 67	2, 43
19	<i>Haloarcula</i> phage SH1	<i>Fuselloviridae</i>	16, 67	4, 51	16, 50	23, 53	5, 60
20	<i>Halo geometricum</i> pleomorphic virus 1	<i>Pleolipoviridae</i>	4, 56	2, 58	1, 55	6, 50	1, 59
21	<i>Halorubrum</i> phage HF2	<i>Pleolipoviridae</i>	11, 55	8, 56	7, 67	32, 60	5, 50
22	<i>Halorubrum</i> pleomorphic virus 1	<i>Pleolipoviridae</i>	9, 52	4, 54	9, 53	7, 68	6, 55
23	<i>Halorubrum</i> pleomorphic virus 3	<i>Pleolipoviridae</i>	5, 67	5, 56	5, 50	14, 51	1, 56
24	<i>Halovirus</i> HCTV-1	<i>Fuselloviridae</i>	62, 59	29, 58	9, 54	99, 51	5, 57
25	<i>Halovirus</i> HCTV-2	<i>Fuselloviridae</i>	38, 57	16, 67	20, 61	39, 63	2, 67
26	<i>Halovirus</i> HF1	<i>Fuselloviridae</i>	45, 68	22, 69	13, 73	73, 60	4, 72
27	<i>Halovirus</i> HGTV-1	<i>Fuselloviridae</i>	148, 67	64, 53	60, 54	235, 71	1, 61
28	<i>Halovirus</i> HRTV-4	<i>Fuselloviridae</i>	32, 54	12, 59	10, 54	37, 67	1, 52
29	<i>Halovirus</i> HRTV-8	<i>Fuselloviridae</i>	65, 61	13, 64	12, 60	66, 55	3, 51
30	<i>Halovirus</i> HSTV-1	<i>Fuselloviridae</i>	26, 57	8, 56	18, 60	25, 59	6, 45
31	<i>Halovirus</i> HSTV-2	<i>Fuselloviridae</i>	30, 43	10, 58	23, 60	76, 61	4, 63
32	<i>Halovirus</i> HVTV-1	<i>Fuselloviridae</i>	55, 53	24, 57	18, 51	89, 60	9, 52
33	<i>Halovirus</i> PH1	<i>Fuselloviridae</i>	19, 60	1, 55	0	23, 61	1, 62
34	<i>Halovirus</i> VNH-1 genomic sequence	<i>Fuselloviridae</i>	10, 59	2, 56	1, 61	20, 65	7, 56
35	His1 virus	<i>Fuselloviridae</i>	28, 56	15, 61	1, 55	39, 54	9, 60
36	Hyperthermophilic archaeal virus 1	Unknown	28, 55	13, 40	2, 51	50, 53	7, 42
37	Hyperthermophilic archaeal virus 2	Unknown	13, 49	4, 50	5, 55	24, 60	8, 54
38	<i>Metallosphaera</i> turreted icosahedral virus strain MTIV1	Unknown	4, 59	2, 60	1, 61	15, 64	3, 65
39	<i>Methanobacterium</i> phage psiM2	<i>Siphoviridae</i>	30, 61	13, 56	10, 59	46, 62	3, 56
40	<i>Natrialba</i> phage PhiCh1	<i>Myoviridae</i>	30, 60	18, 68	18, 70	53, 63	8, 51
41	<i>Pyrobaculum</i> filamentous virus 1	<i>Globuloviridae</i>	32, 59	13, 61	13, 63	47, 76	6, 44
42	<i>Pyrobaculum</i> spherical virus	<i>Globuloviridae</i>	22, 55	15, 51	12, 49	53, 61	8, 45
43	<i>Pyrococcus abyssi</i> virus 1	<i>Rudiviridae</i>	24, 60	11, 55	10, 61	28, 59	8, 63
44	<i>Sulfolobales</i> Mexican fusellovirus 1	<i>Fuselloviridae</i>	19, 56	6, 67	3, 55	33, 49	9, 45
45	<i>Sulfolobales</i> Mexican rudivirus 1	<i>Fuselloviridae</i>	28, 71	18, 60	13, 63	61, 67	4, 51
46	<i>Sulfolobales</i> virus YNP1 strain SYV1	<i>Fuselloviridae</i>	70, 52	29, 67	40, 54	137, 61	8, 50
47	<i>Sulfolobales</i> Virus YNP2 strain SYV2	<i>Fuselloviridae</i>	54, 58	14, 60	15, 67	75, 66	9, 79
48	<i>Sulfolobus islandicus</i> filamentous virus	<i>Rudiviridae</i>	111, 55	47, 61	89, 51	197, 59	4, 41

(Continued)

Table 3 | Continued

S. No	Best blast hit (BBH) annotation of contigs	Archaeal virus family	MNS1 (# of BBH, average %identity)	MNS2 (# of BBH, average %identity)	MNS3 (# of BBH, average %identity)	MNS4 (# of BBH, average %identity)	Mat (# of BBH, average %identity)
49	<i>Sulfolobus islandicus</i> rod-shaped virus 10, partial genome	<i>Rudiviridae</i>	0	101, 67	10, 51	351, 49	6, 60
50	<i>Sulfolobus islandicus</i> rod-shaped virus 2	<i>Rudiviridae</i>	219, 59	84, 60	78, 55	324, 63	5, 61
51	<i>Sulfolobus islandicus</i> rudivirus 3 isolate SIRV3	<i>Rudiviridae</i>	196, 55	116, 58	120, 60	313, 61	5, 59
52	<i>Sulfolobus monocaudavirus</i> SMV1 complete genome	<i>Rudiviridae</i>	0	40, 61	28, 59	113, 64	3, 61
53	<i>Sulfolobus monocaudavirus</i> SMV4	<i>Rudiviridae</i>	82, 55	38, 49	18, 52	144, 51	1, 56
54	<i>Sulfolobus</i> turreted icosahedral virus	<i>Rudiviridae</i>	49, 55	21, 56	19, 54	75, 55	7, 54
55	<i>Sulfolobus</i> virus 1 complete genome (provirus)	<i>Rudiviridae</i>	0	20, 67	19, 71	56, 55	9, 49
56	<i>Sulfolobus</i> virus Kamchatka 1	<i>Rudiviridae</i>	38, 61	15, 65	20, 68	56, 62	0
57	<i>Sulfolobus</i> virus Ragged Hills	<i>Rudiviridae</i>	38, 61	21, 71	12, 73	55, 60	0
58	<i>Sulfolobus</i> virus STSV2	<i>Rudiviridae</i>	198, 59	96, 58	70, 55	348, 57	3, 58
59	<i>Thermococcus prieurii</i> virus 1	<i>Fuselloviridae</i>	27, 50	8, 49	18, 55	48, 56	8, 59

The BLAST output was parsed for Best Blast Hits (BBH) based on *E*-value, bit score, and coverage cut-off of 80%. The BBH annotations for the metagenome contigs that showed a match are mentioned in the table along with % identity (average) and number of BBH hits for each archaeal virus type (i.e., number of contigs showing a match to the archaeal virus).

portal protein, integrase, baseplate assembly proteins, and phage related replication machinery. Notably, the phage genomes reconstructed from both sediment and microbial mat samples also carried a common core encoding for non-phage proteins, such as hybrid sensor histidine kinase, an arsenic resistance operon repressor, Clp protease, hydroxypyruvate reductase, and methyltransferase. The presence of integrase genes across both sample types suggests that lysogeny is common in thermal discharges, which is consistent with previous studies (Schoenfeld et al., 2008). Additionally, the bacterial genes encoding Clp protease, DNA methyltransferases, and histidine kinase in the phage genomes have been implicated in regulation of protein synthesis, removal of misfolded proteins, antibiotic resistance and virulence in bacteria (Krüger et al., 2000; Pedersen et al., 2010; Tang et al., 2013; Malik and Brötz-Oesterhelt, 2017). These results suggest a role for viral transmission of these functions in bacterial fitness via lateral gene transfers.

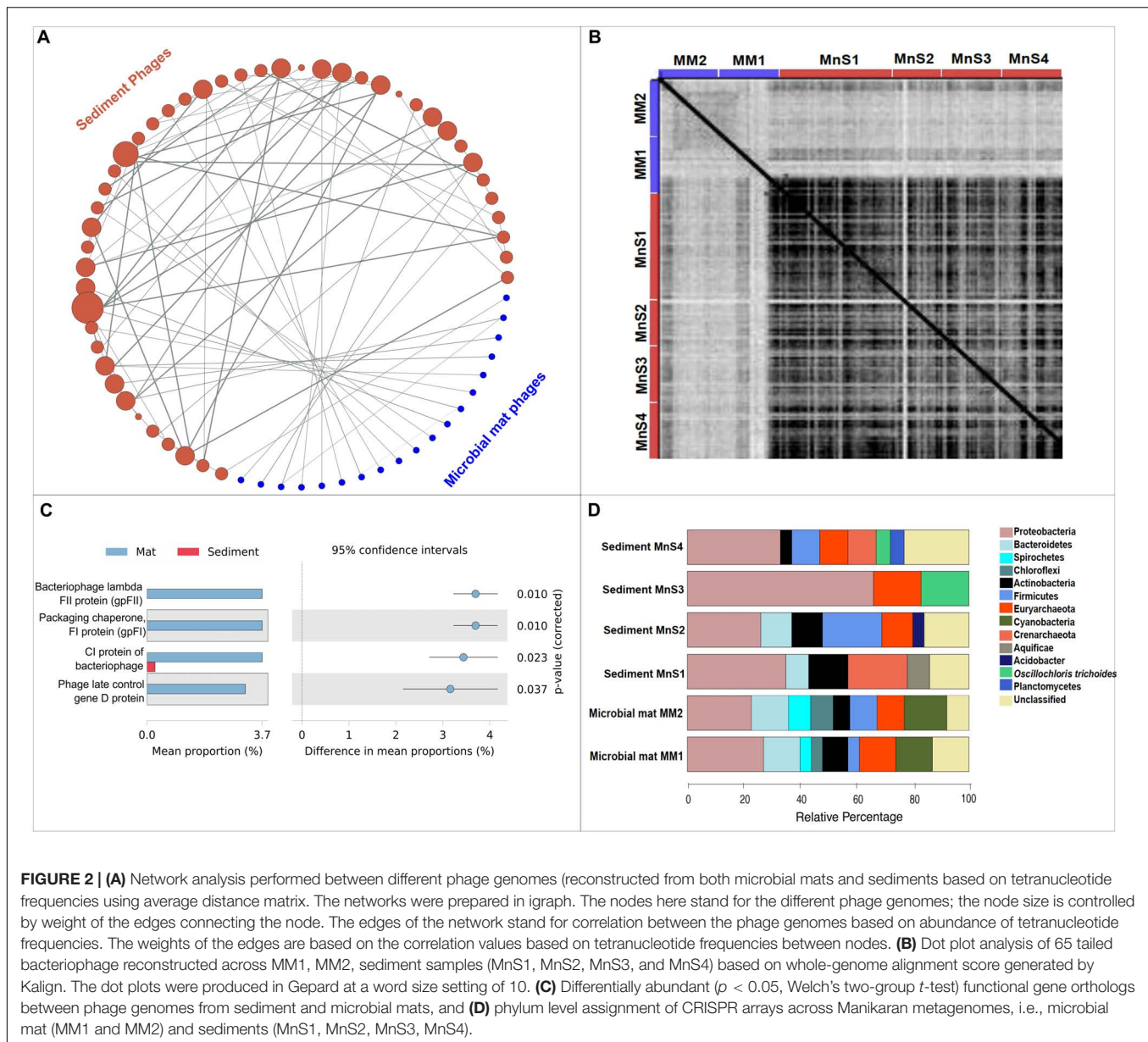
Site-specific phage orthologous proteins were identified in all sediment (MnS1–MnS4) and microbial mat samples (MM1, MM2) so that the influence of micro-niches on the genetic repertoire of the phage community could be determined. We again identified presence of both phage and non-phage proteins among the set of shared orthologous proteins at both sites. The orthologous proteins ( $n = 28$ ) between the genomes of phages reconstructed from microbial mat samples, included transporter units, SbcC phage protein, and RnhA endonuclease protein. The reconstructed sediment phages also shared common orthologs ( $n = 49$ ) including both phage and non-phage genes. Besides the phage-related replicative and structural machinery, the genomes contained regions coding for TonB-dependent receptors, histidine kinase, ClpC protease, thymidylate synthase, transposase, integrase, etc. TonB-dependent receptors play a role in infection for *Siphoviridae* phages (Rabsch et al., 2007). Histidine kinase was recently reported in phage genomes for the first time and implicated directly in bacterial biofilm formation

and also pathogenicity (Hargreaves et al., 2014). The presence of these genes, along with phage-related genes, integrases, and transposases, among orthologs suggest potential lateral gene transfer activity between these phage genomes and their hosts in this extreme environment.

## Metagenomic Reconstruction and Motif Level Assignment of CRISPR Cassettes

Since viruses are predatory in nature, it is unsurprising that both bacterial and archaeal hosts have developed a countermeasure—the CRISPR system—which confers adaptive immunity against future viral attacks (Rath et al., 2015). The presence of intact CRISPR arrays provide molecular evidence of prior viral infections and resulting adaptive immunity. CRISPR cassettes were predicted using both PILER-CR and CRISPRFinder, with the consistent arrays retained in the hopes of minimizing the false positives. The final set of CRISPRs consisted of 7, 45, and 455 cassettes from MM2, MM1, and sediment (MnS1 = 307, MnS2 = 54, MnS3 = 17, MnS4 = 77), respectively. CRISPR cassettes were found to be composed of copies of a ~35-nucleotide (nt) direct repeat (DR) region separated by ~34-nt long spacer sequences. Each repeat sequence of a CRISPR cassette is known to be unique and thus is used to characterize the CRISPR type (Skennerton et al., 2013). Among the total set of 507 CRISPR arrays, 4, 23, 14, 19, 6, and 21 different CRISPR cassettes with unique repeat type were found across MM2, MM1, and sediments – MnS1, MnS2, MnS3, and MnS4, respectively (see **Supplementary Table S2**). Out of 87 distinct CRISPR types, the majority (35.6%) was assigned to phylum Proteobacteria, followed by Actinobacteria (8%), Firmicutes (8%), Euryarchaeota (8%), and Bacteroidetes (6.9%) (**Figure 2D**). Five CRISPR cassettes were related to Crenarchaeota (5.7%) and three each from Cyanobacteria and Chloroflexi (**Figure 2D**). Only two CRISPR cassettes were assigned a species level designation of





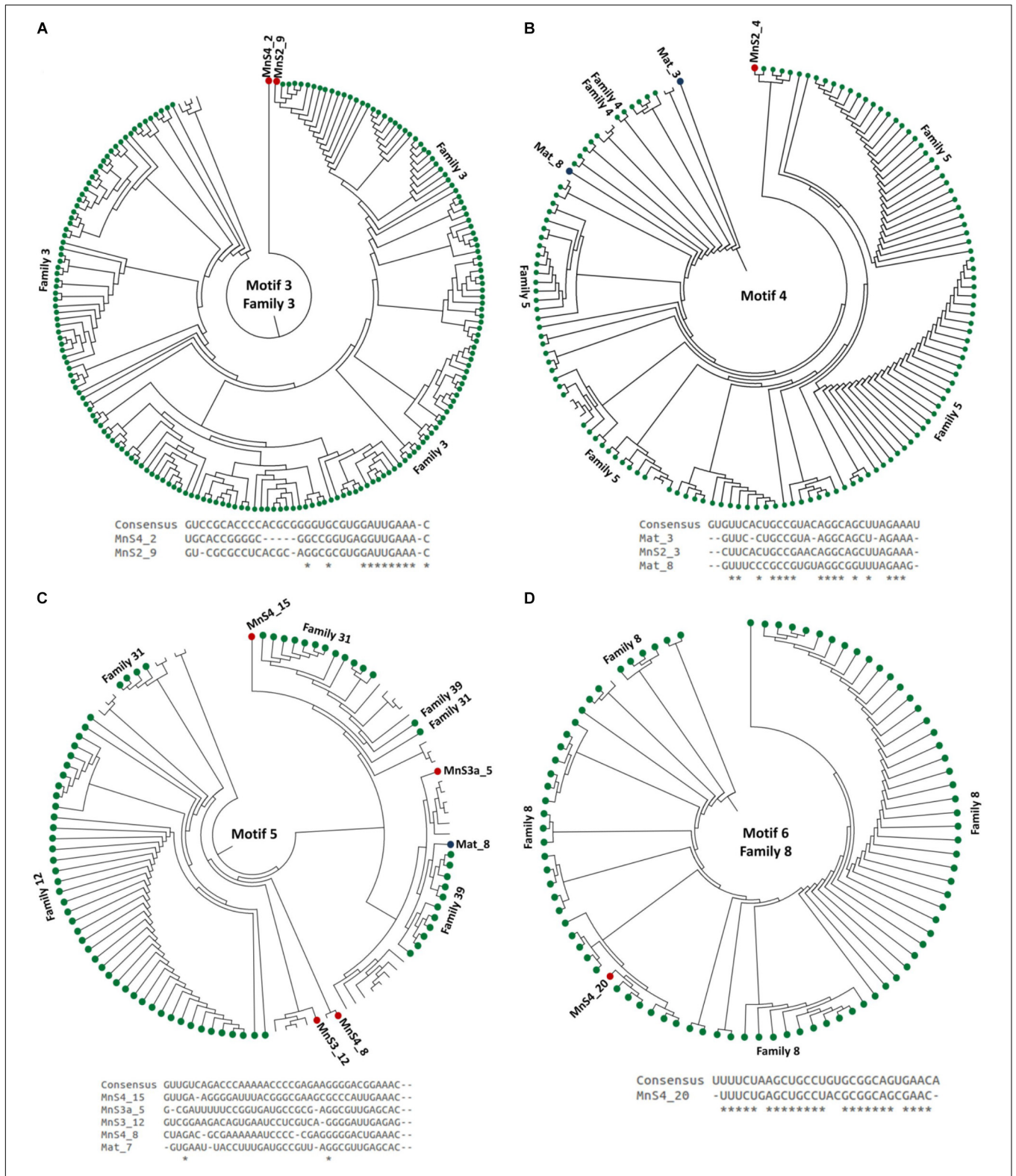
*Oscillochloris trichoides* (Figure 2D). This analysis also suggests that there is an uneven distribution of CRISPRs across the prokaryotic phyla. Recent explanations of this phenomena point to the predominance of symbiotic lifestyles in bacteria, or the potential that the maintenance costs a CRISPR systems outweigh the conferred defensive benefits due to frequent auto-immunity (Burstein et al., 2016).

After characterizing different CRISPR types based on unique repeat sequences, we classified these unique CRISPRs based on key binding motif (present within the repeats) for Cas endoribonucleases, which play a significant role in the adaptive immunity conferred by CRISPR cassettes (Lange et al., 2013). Cas endoribonucleases bind to CRISPR repeat sequences for identification and inhibition of attacks by phages. The binding affinity of Cas proteins is dependent on a key binding motif

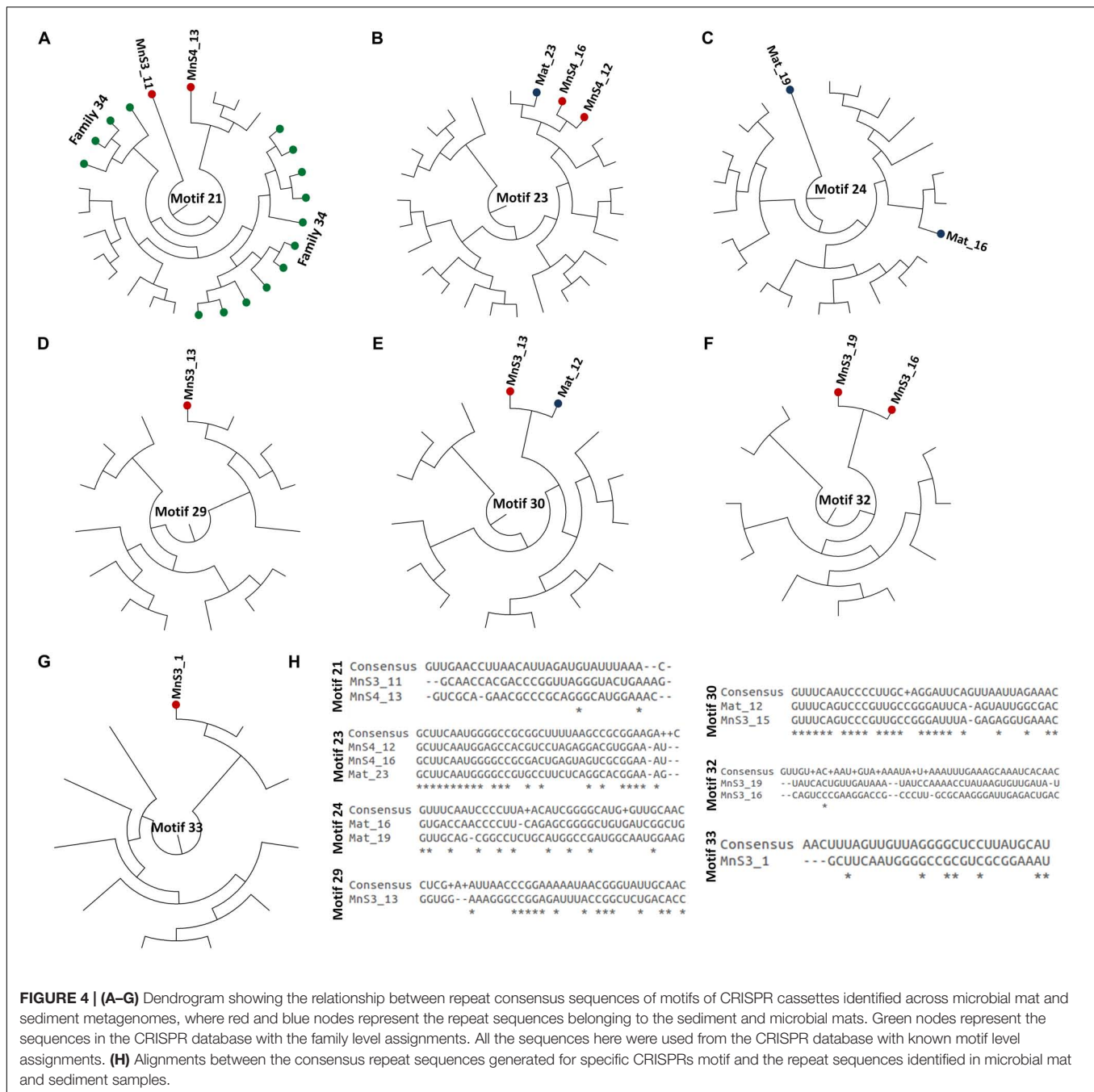
present in the repeat sequence, and here we have characterized the reconstructed CRISPR cassettes based on these motifs 1–33 (Lange et al., 2013). Out of 87 different CRISPR cassettes identified across both microbial mat and sediment metagenomes, only 24 were assigned with these motif types—just 28% of the repeats showed matches in the existing database, with the 72% of CRISPR arrays being novel (Figures 3, 4). These results again highlight the limited database that exists for viruses overall and how much is still unknown and uncharacterized.

## Metagenomic Reconstruction of Potential Bacteriophage Hosts

Bacterial diversity analyses (based on essential genes identified using MetaPhlan) of sediment and microbial mat samples

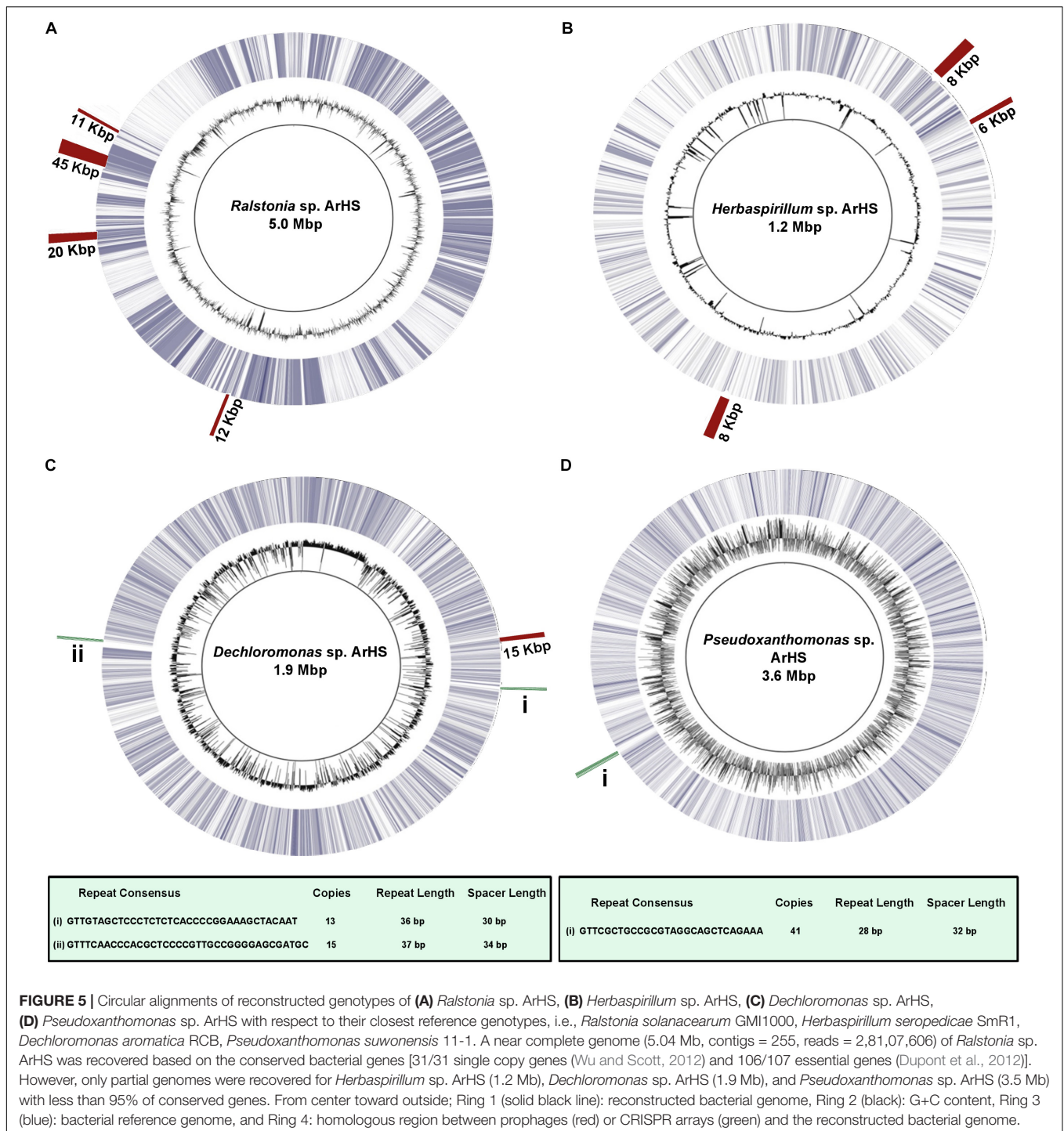


**FIGURE 3 |** Dendrograms showing the relationship between repeat consensus sequences of the CRISPR cassettes belonging to (A) motif 1, (B) Motif 2, (C) Motif 3, and (D) Motif 4, identified across microbial mat and sediment metagenomes. The red and blue nodes represent the repeat sequences belonging to the sediment and microbial mats. Green nodes represent the sequences in the CRISPR database with the family level assignments. All the sequences here were used from the CRISPR database with known motif level assignments. Below each dendrogram, there is an alignment between the consensus repeat sequences generated for a specific CRISPRs motif and the repeat sequences identified in microbial mat and sediment samples.



revealed an abundance of genera like *Ralstonia*, *Burkholderia*, *Enterobacteria*, *Herbaspirillum*, etc. Draft genomes of four potential phage hosts, i.e., *Ralstonia* (5 Mb), *Pseudoxanthomonas* (3.5 Mb), *Dechloromonas* (1.9 Mb), and *Herbaspirillum* (1.2 Mb) were reconstructed from microbial mats (MM2) (Figure 5). Metagenomic fragment recruitment of MM2 samples also revealed significant coverage for the reference genomes *Dechloromonas aromatica* RCB ( $n = 32,64,075$ ), *Herbaspirillum seropedicae* SmR1 ( $n = 71,57,193$ ), *Pseudoxanthomonas suwonensis* 11-1 ( $n = 35,05,890$ ), and *Ralstonia mannitolilytica* ( $n = 70,30,087$ ) (Supplementary Figure S4). A near complete

genome (5.04 Mb, contigs = 255, reads = 2,81,07,606) of *Ralstonia* sp. ArHS was recovered and further characterized ( $P$ -value < 0.001) at the genus level by MetaBin (relative abundance, 8.5%) (Sharma et al., 2012). *Ralstonia* sp. ArHS was annotated with 4,819 coding sequences. Genome-wide analysis revealed the presence of four prophages with sizes 12.2 kb, 20.7 kb, 44.8 kb, and 11.2 kb (Figure 5 and Supplementary Figure S5). Interestingly, the CRISPR-associated gene *cas1* (which is generally present upstream or downstream of CRISPR array) was found in the genome of strain ArHS, however, we were not able to identify intact CRISPR array in



the genome. The identification of prophages in the genome suggests that *Ralstonia* sp. ArHS is a phage host in the microbial mats of the Manikaran hot springs (Rath et al., 2015).

Partial genomes of *Dechloromonas* sp. ArHS (1.9 Mb) and *Herbaspirillum* sp. ArHS (1.2 Mb) were screened for prophages and CRISPR arrays. One prophage (13.5 kb) was identified in *Dechloromonas* sp. ArHS and three prophages (8 kb, 6 kb,

and 8 kb) were identified in *Herbaspirillum* sp. ArHS. Further, composition-based analysis of these arrays revealed two distinct types (based on repeat sequences) in *Dechloromonas* sp. ArHS, with 13 and 15 copies across the genome (Supplementary Table S3 and Supplementary Figures S6, S7). Similarly, *Pseudoxanthomonas* sp. ArHS showed 41 copies of an array with unique repeat type (Supplementary Figure S7). The presence of intact CRISPR arrays in the reconstructed genotypes

of *Dechloromonas* and *Pseudoxanthomonas*, specifically short fragments-spacers, suggests prior frequent phage infection and acquired immunity against the same phage over time.

## CONCLUSION

We employed metagenomics and electron-ion microscopy to study viral diversity at the Manikaran hot springs in India. We investigated viral enrichment cultures produced from sediment and microbial mat samples, which revealed the presence of archaeal viruses belonging to family *Fuselloviridae* with varying sizes within sediments, and tailed or tailless phages within microbial mats. Further, we used metagenomic data from microbial mat ( $n = 2$ ) and sediment ( $n = 4$ ) samples to reconstruct 65 bacteriophage genomes and 4 potential phage hosts. Combined, these genomes reveal potential predator–host interactions at the Manikaran hot springs using functional annotations of both viral and bacterial genomes. Both sediment and microbial mat phage genomes showed presence of both phage and host-associated genes suggesting potential lateral gene transfer activity between these phage genomes and their hosts in this extreme environment. This study has elucidated the prokaryotic virosphere at the Himalayan hot springs of Manikaran, providing a glimpse of the ecological roles and community dynamics of viruses, bacteria and archaea in geothermal environments. The genetic information gained in this study can be used to guide microscopy investigations (for example, with FISH probes), as well as direct isolation and cultivation efforts of Manikaran ecotypes. We provide all the data presented here as a resource for future virome studies of extreme thermophilic environments.

## MATERIALS AND METHODS

### Sample Collection and Metagenome Assembly

Microbial mat and sediment samples were collected from Manikaran hot springs on 27th October, 2014 (see **Supplementary Figure S1**). Metagenomic DNA was isolated according to Sangwan et al., 2015). Paired-end reads were generated using an Illumina HiSeq 2000 ( $n = 42,418,084$ ; 2 kb paired-end library) at the Beijing Genome Institute, BGI, Shenzhen, Guangdong, China. For comparative metagenomic analysis, we used data from a previously sequenced microbial mat (named MM1) (Sangwan et al., 2015) as a baseline of comparison for the four sediment (MnS1, MnS2, MnS3, and MnS4) and one microbial mat sample (MM2) sequenced in this study.

### Reconstruction of Phage Genomes and Phage Comparison

Phages were recovered from the metagenome data using the approach described in Smits et al. (2015), beginning with

identification of the viral load by BLASTX against phage databases<sup>1</sup> (Smits et al., 2015). The contigs were initially binned on the basis of oligonucleotide composition, contig length, and %G+C content (Sangwan et al., 2015). Phage assembly is challenging due to a greater diversity than that of bacterial genomes, thus we applied an iterative assembly strategy using PRICE assembler to overcome the coverage bias (Ruby et al., 2013). After achieving nearly complete viral genomes, *k*-mer profiling (tetranucleotide frequencies) was used to link the fragments together and allow the draft optimal genome recovery (Smits et al., 2015). In order to annotate the phages, ORFs were predicted using Prodigal (Hyatt et al., 2010) followed by BLASTP against NCBI non-redundant protein collection (NR) (at E-value  $1e-5$ ). Additionally, phage genomes were checked for functional completeness by calculating the functional completeness score employed in PHAST-Enhanced Release which uses over 200,000 annotated phage genomes as reference (Zhou et al., 2011; Arndt et al., 2016). The completeness score is calculated based on (i) identification of cornerstone genes, which include key structural genes (i.e., capsid, head, tail, coat, portal, and holin), phage DNA regulation genes (such as integrase, transposase, terminase), and phage function genes (such as lysin, bacteriocin) (Casjens, 2003), and (ii) identification of genes present in already known phage in reference database (refer to Zhou et al., 2011 and Arndt et al., 2016 for details on completeness score calculation criteria). Phage genomes were compared using a dot-plot based on pairwise whole genome sequences in Gepard\_1.30 program (Krumisiek et al., 2007). Core genes were calculated between sediment and microbial mat metagenomes by using the pairwise reciprocal smallest distance (RSD) algorithm at an *E*-value and divergence cut-off of  $1e-15$  and 0.5, respectively (Wall and Deluca, 2007). Phage genomes were analyzed for the lifestyle [i.e., virulent (lytic) or temperate] using Phage Classification Tool Set (PHACTS) version 0.3 program employing a supervised Random Forest algorithm for classification of phage ORFs with a probability cut off value of 0.05 (McNair et al., 2012). In order to assign functionality to the phage metagenomes, we performed a HMMER 3.1b (Finn et al., 2011) search against the Caudovirales orthologous gene HMM profile<sup>2</sup> downloaded from the EggNOG 4.5 database<sup>3</sup>.

In order to screen mat and sediment samples for archaeal viruses, we downloaded complete genomes of archaeal viruses ( $n = 95$ ) from NCBI<sup>4</sup>. BLASTX was used to screen the metagenome contigs against the archaeal database. Using Bioperl script (NCBI Blast Parser<sup>5</sup>), the BLASTX output was then parsed to retrieve the Best BLAST Hit (BBH) for each contig based on E-value, bit score and coverage cut-off of 80%.

<sup>1</sup><http://www.phantome.org/Downloads/>

<sup>2</sup>[http://eggnogdb.embl.de/download/eggnog\\_4.5/data/viruses/Caudovirales/](http://eggnogdb.embl.de/download/eggnog_4.5/data/viruses/Caudovirales/)

<sup>3</sup><http://eggnogdb.embl.de/#/app/viruses>

<sup>4</sup><https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239>

<sup>5</sup><http://www.bioinformatics-made-simple.com/2012/07/ncbi-blast-parser-extract-query-and.html>

## Helium-Ion Microscopy Analysis of Sediment and Mat Samples for Bacterial Diversity

Microbial mat and sediment samples were surveyed under high-resolution scanning HIM in order to investigate the bacterial diversity in each sample. In preparation of microscopic analysis, samples were dehydrated by increasing ethanol concentration steadily to 70% within an hour. At this point, the samples were stored in the fridge at 4°C for 24 h so as to exploit the gentle fixative effect of ethanol on cell walls. After 1 day, the ethanol concentration was increased to 100%. Next, the samples were dried in a fully automated Leica CPD300 (Austria) critical point dryer, after which the samples were mounted on standard SEM stubs using a silver-epoxy.

The HIM was carried out with a Zeiss (United States) ORION nanoFab helium-ion microscope. The ion energy and beam current amounted to 25 kV and 0.6 pA, respectively. For the imaging, we employed the Everhard-Thornley secondary-electron detector.

## Virus Enrichment Followed by Electron Microscopy

A mass of 5 g from each of the microbial mat and sediments samples was deposited into 50 mL falcons and 0.1 M ammonium acetate was added until a total volume of 45 mL was reached. The samples were then placed on a shaker for 2 h and kept at room temperature in an incubator. Following that, the tubes were centrifuged at  $10,000 \times g$  for 20 min at 4°C. The supernatant was filtered to remove cells and cell debris using 0.22  $\mu\text{m}$  syringe filters (Merck Millipore, Germany). The filtrate was stored at 4°C until downstream processing. Virus concentration and purification was carried out with two aliquots (5.5 mL in polyallomer tubes) of each sample in a table-top ultracentrifuge Optima MAX-XP (Beckman Coulter, United States) equipped with a swinging bucket rotor MLS-50 (Beckman Coulter, United States). The samples were spun at  $80,000 \times g$  at 4°C under vacuum conditions for 4 h. The supernatant was discarded and 50  $\mu\text{L}$  of 0.1 M ammonium acetate (pH 7.0) was added to each centrifuge tube to re-suspend the virus pellet and left overnight.

Afterward, 10  $\mu\text{L}$  aliquots of the virus preparations were placed onto parafilm, onto which a 300-mesh formvar-coated copper grid (Plano, Germany) was placed (film down) and left to sit for 1 min in order to allow the viruses to attach to the grid. The excess liquid was blotted off the surface of the grid, and the dried grid was stained twice with a 2% uranium acetate solution, for 30 s each time. The grid was left to dry in a petri-dish overnight, after which the samples were analyzed using a Zeiss (Germany) Merlin VP compact SEM with an option of transmission imaging.

## Environmental CRISPR Identification and Characterization

Clustered Regularly Interspaced Short Palindromic Repeats were identified in the metagenome samples by using PILER-CR v1.06 (Edgar, 2007) and CRISPRFinder (Grissa et al., 2007). CRISPR arrays were assigned taxonomic status by using contigs and

CRISPR repeats through a BLASTx search against NCBI non-redundant protein collection (NR) with an E-value  $10^{-5}$ . For both approaches, taxonomic status was assigned if the top 10 hits belonged to one phylum; taxonomic labels at the level of class, family, and genus were assigned if the majority of the top 30 hits belonged to the same taxon of that level. If the top 10 labels for that specific contig or repeat were relatively diverse, then no status was assigned. Moreover, the CRISPR cassettes were characterized at the motif (1–33) and family (1–40) levels based on a repeat consensus using CRISPRMap (Lange et al., 2013).

## Taxonomic Binning for Potential Phage Host Identification

Microbial mat ( $n = 2$ ) and sediment ( $n = 4$ ) metagenomes were binned for phylogenetic composition using MetaBin (Sharma et al., 2012), MetaPhlan (Segata et al., 2012), and BLASTx bit-score with a Lowest Common Ancestor (LCA) approach. The contigs (>10 kb) were clustered in R4.3.2 on the basis of tetranucleotide frequency, coverage, and %G+C content to generate multiple bins at each parameter<sup>6</sup>. Broadly, four distinct bins were generated based on tetranucleotide frequency and coverage, to which they were further binned for taxonomical assignments and revealed four independent genera—*Ralstonia* (5.1 Mb), *Pseudoxanthomonas* (3.5 Mb), *Dechloromonas* (1.9 Mb), and *Herbaspirillum* (1.2 Mb). Finally, each bin was then analyzed for completeness through validation using 107 essential (Segata et al., 2012) and 31 bacterial marker genes (Wu and Scott, 2012) leading to a result of 99.9% completeness for *Ralstonia* sp., which was then labeled *Ralstonia* sp. ArHS.

## AVAILABILITY OF DATA AND MATERIALS

Metagenome sequence data for microbial mats was deposited under accession numbers: MM1-PRJEB4614 (<http://www.ebi.ac.uk/ena/data/view/PRJEB4614>) and MM2-PRJEB14856 (<http://www.ebi.ac.uk/ena/data/view/PRJEB14856>) in EBI. All the sediment samples (MnS1, MnS2, MnS3, and MnS4) were submitted to DDBJ/EMBL/GenBank under the accession number: PRJEB14843 (<http://www.ebi.ac.uk/ena/data/view/PRJEB14843>), respectively. The metagenome data for MM2 were submitted under the accession number PRJEB14856. The phage genomes recovered from microbial mats and sediments were deposited under the accession numbers ERS2768334-ERS278398 under the project PRJEB14843.

## AUTHOR CONTRIBUTIONS

RL developed the concept for the study. RL and AS performed the metagenomic analyses. AS, MS, BK, and NM did the viral enrichments and analyzed the samples under ion/electron microscopes. RL, AS, NM, LC, YS, HR, WA, and JG wrote the manuscript.

<sup>6</sup><http://www.R-project.org/>

## ACKNOWLEDGMENTS

The authors acknowledge funds from the Department of Biotechnology (DBT), Government of India, National Bureau of Agriculturally Important Microorganisms (NBAIM), DUDST Promotion of University Research and Scientific Excellence (PURSE) grant, and German Academic Exchange Services (DAAD). AS and NM gratefully acknowledge NBAIM, DAAD, and DBT for providing research fellowships. The phage enrichment and electron-ion microscopy were performed at Department of Isotope Biogeochemistry at Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany, during AS's visit under DAAD-sponsored Bi-National Doctoral Degree

Program. The authors are grateful for the use of the analytical facilities at the Centre for Chemical Microscopy (ProVIS), a part of the Helmholtz Centre for Environmental Research, Leipzig, which is supported by European Regional Development Funds (EFRE – Europe funds Saxony) and the Helmholtz Association.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.03095/full#supplementary-material>

## REFERENCES

- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387
- Breitbart, M., Miyake, J. H., and Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* 236, 249–256. doi: 10.1016/j.femsle.2004.05.042
- Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613
- Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* 49, 277–300. doi: 10.1046/j.1365-2958.2003.03580.x
- Clokic, M. R., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011). Phages in nature. *Bacteriophage* 1, 31–45. doi: 10.4161/bact.1.1.14942
- Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Richter, R. A., Valas, R., et al. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6, 1186–1199. doi: 10.1038/ismej.2011.189
- Dwivedi, V., Sangwan, N., Nigam, A., Garg, N., Niharika, N., Khurana, P., et al. (2012). Draft genome sequence of thermus sp. strain RL, isolated from a hot water spring located atop the himalayan ranges at manikaran, India. *J. Bacteriol.* 194, 3534–3534. doi: 10.1128/JB.00604-12
- Edgar, R. C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* 8:18. doi: 10.1186/1471-2105-8-18
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360
- Hargreaves, K. R., Kropinski, A. M., and Clokic, M. R. J. (2014). What does the talking?: quorum sensing signalling genes discovered in a bacteriophage genome. *PLoS One* 9:e85131. doi: 10.1371/journal.pone.0085131
- Häring, M., Vestergaard, G., Brügger, K., Rachel, R., Garrett, R. A., and Prangishvili, D. (2005). Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. *J. Bacteriol.* 187, 3855–3858. doi: 10.1128/JB.187.11.3855-3858.2005
- Hatfull, G. F., Jacobs-Sera, D., Lawrence, J. G., Pope, W. H., Russell, D. A., Ko, C.-C., et al. (2010). Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J. Mol. Biol.* 397, 119–143. doi: 10.1016/j.jmb.2010.01.011
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Krüger, E., Witt, E., Ohlmeier, S., Hanschke, R., and Hecker, M. (2000). The Clp Proteases of *Bacillus subtilis* are directly involved in degradation of misfolded proteins. *J. Bacteriol.* 182, 3259–3265. doi: 10.1128/JB.182.11.3259-3265.2000
- Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026–1028. doi: 10.1093/bioinformatics/btm039
- Krupovic, M., Prangishvili, D., Hendrix, R. W., and Bamford, D. H. (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.* 75, 610–635. doi: 10.1128/MMBR.00111-11
- Lange, S. J., Alkhnbashi, O. S., Rose, D., Will, S., and Backofen, R. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* 41, 8034–8044. doi: 10.1093/nar/gkt606
- Mahato, N. K., Tripathi, C., Verma, H., Singh, N., and Lal, R. (2014). Draft genome sequence of *Deinococcus* sp. strain RL isolated from sediments of a hot water spring. *Genome Announc.* 2, e703–e714. doi: 10.1128/genomeA.00703-14
- Malik, I., and Brötz-Oesterhelt, H. (2017). Conformational control of the bacterial Clp protease by natural product antibiotics. *Nat. Prod. Rep.* 34, 815–831. doi: 10.1039/C6NP00125D
- McNair, K., Bailey, B. A., and Edwards, R. A. (2012). PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinformatics* 28, 614–618. doi: 10.1093/bioinformatics/bts014
- Munson-McGee, J. H., Snyder, J. C., and Young, M. J. (2018). Archaeal viruses from high-temperature environments. *Genes* 9:E128. doi: 10.3390/genes9030128
- Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* 32, 822–828. doi: 10.1038/nbt.2939
- Pedersen, M., Ligowska, M., and Hammer, K. (2010). Characterization of the CI repressor protein encoded by the temperate Lactococcal phage TP901-1. *J. Bacteriol.* 192, 2102–2110. doi: 10.1128/JB.01387-09
- Pietilä, M. K., Demina, T. A., Atanasova, N. S., Oksanen, H. M., and Bamford, D. H. (2014). Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol.* 22, 334–344. doi: 10.1016/j.tim.2014.02.007
- Prangishvili, D. (2013). The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.* 67, 565–585. doi: 10.1146/annurev-micro-092412-155633
- Prangishvili, D., Koonin, E. V., and Krupovic, M. (2013). Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in archaea. *Biochem. Soc. Trans.* 41, 443–450. doi: 10.1042/BST20120313
- Rabsch, W., Ma, L., Wiley, G., Najjar, F. Z., Kaserer, W., Schuerch, D. W., et al. (2007). FepA- and TonB-dependent bacteriophage H8: receptor binding and genomic sequence. *J. Bacteriol.* 189, 5658–5674. doi: 10.1128/JB.00437-07
- Rachel, R., Bettstetter, M., Hedlund, B. P., Häring, M., Kessler, A., Stetter, K. O., et al. (2002). Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch. Virol.* 147, 2419–2429. doi: 10.1007/s00705-002-0895-2
- Rath, D., Amlinger, L., Rath, A., and Lundgren, M. (2015). The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* 117, 119–128. doi: 10.1016/j.biochi.2015.03.025

- Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001). Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13341–13345. doi: 10.1073/pnas.231170198
- Rohwer, F., and Edwards, R. (2002). The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol.* 184, 4529–4535. doi: 10.1128/JB.184.16.4529-4535.2002
- Ruby, J. G., Bellare, P., and Derisi, J. L. (2013). PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3* 3, 865–880. doi: 10.1534/g3.113.005967
- Sangwan, N., Lambert, C., Sharma, A., Gupta, V., Khurana, P., Khurana, J. P., et al. (2015). Arsenic rich himalayan hot spring metagenomics reveal genetically novel predator-prey genotypes. *Environ. Microbiol. Rep.* 7, 812–823. doi: 10.1111/1758-2229.12297
- Sangwan, N., Xia, F., and Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome* 4:8. doi: 10.1186/s40168-016-0154-5
- Schoenfeld, T., Patterson, M., Richardson, P. M., Wommack, K. E., Young, M., and Mead, D. (2008). Assembly of viral metagenomes from yellowstone hot springs. *Appl. Environ. Microbiol.* 74, 4164–4174. doi: 10.1128/AEM.02598-07
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066
- Sharma, A., Gilbert, J. A., and Lal, R. (2016a). (Meta)genomic insights into the pathogenome of *Cellulosimicrobium cellulans*. *Sci. Rep.* 6:25527. doi: 10.1038/srep25527
- Sharma, A., Kohli, P., Singh, Y., Schumann, P., and Lal, R. (2016b). *Fictibacillus halophilus* sp. nov., from a microbial mat of a hot spring atop the himalayan range. *Int. J. Syst. Evol. Microbiol.* 66, 2409–2416. doi: 10.1099/ijsem.0.001051
- Sharma, A., Hira, P., Shakarad, M., and Lal, R. (2014). Draft genome sequence of *Cellulosimicrobium* sp. strain MM, isolated from arsenic-rich microbial mats of a himalayan hot spring. *Genome Announc.* 2:e01020-14. doi: 10.1128/genomeA.01020-14
- Sharma, V. K., Kumar, N., Prakash, T., and Taylor, T. D. (2012). Fast and accurate taxonomic assignments of metagenomic sequences using MetaBin. *PLoS One* 7:e34030. doi: 10.1371/journal.pone.0034030
- Skenneron, C. T., Imelfort, M., and Tyson, G. W. (2013). Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res.* 41:e105. doi: 10.1093/nar/gkt183
- Smits, S. L., Bodewes, R., Ruiz-González, A., Baumgärtner, W., Koopmans, M. P., Osterhaus, A. D. M. E., et al. (2015). Recovering full-length viral genomes from metagenomes. *Front. Microbiol.* 6:1069. doi: 10.3389/fmicb.2015.01069
- Tang, F., Bossers, A., Harders, F., Lu, C., and Smith, H. (2013). Comparative genomic analysis of twelve *Streptococcus suis* (pro)phages. *Genomics* 101, 336–344. doi: 10.1016/j.ygeno.2013.04.005
- Tripathi, C., Mahato, N. K., Singh, A. K., Kamra, K., Korpole, S., and Lal, R. (2016). *Lampropedia cohaerens* sp. nov., a biofilm-forming bacterium isolated from microbial mats of a hot water spring, and emended description of the genus *Lampropedia*. *Int. J. Syst. Evol. Microbiol.* 66, 1156–1162. doi: 10.1099/ijsem.0.000853
- Wagner, C., Reddy, V., Asturias, F., Khoshouei, M., Johnson, J. E., Manrique, P., et al. (2017). Isolation and characterization of *Metallosphaera* turreted icosahedral virus (MTIV), a founding member of a new family of archaeal viruses. *J. Virol.* doi: 10.1128/JVI.00925-17 [Epub ahead of print].
- Wall, D. P., and Deluca, T. (2007). Ortholog detection using the reciprocal smallest distance algorithm. *Methods Mol. Biol.* 396, 95–110. doi: 10.1007/978-1-59745-515-2\_7
- Wu, M., and Scott, A. J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28, 1033–1034. doi: 10.1093/bioinformatics/bts079
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Sharma, Schmidt, Kiesel, Mahato, Cralle, Singh, Richnow, Gilbert, Arnold and Lal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.