Applied and Environmental Microbiology, Nov. 2010, p. 7251–7258 Vol. 76, No. 21
0099-2240/10/$12.00 doi:10.1128/AEM.01109-10
Copyright © 2010, American Society for Microbiology. All Rights Reserved.

Use of Cellular CRISPR (Clusters of Regularly Interspaced Short Palindromic Repeats) Spacer-Based Microarrays for Detection of Viruses in Environmental Samples▽†

Jamie C. Snyder,[1,2] Mary M. Bateson,[1,2] Matthew Lavin,[2] and Mark J. Young[1,2]*

Thermal Biology Institute[1] and Department of Plant Sciences and Plant Pathology,[2] Montana State University, Bozeman, Montana 59717

Received 7 May 2010/Accepted 23 August 2010

It is currently difficult to detect unknown viruses in any given environment. The recent discovery of CRISPR (clusters of regularly interspaced short palindromic repeats) loci within bacterial and archaeal cellular genomes may provide an alternative approach to detect new viruses. It has been shown that the spacer sequences between the direct repeat units of the CRISPR loci are often derived from viruses and likely function as guide sequences to protect the cell from viral infection. The spacer sequences within the CRISPR loci may therefore serve as a record of the viruses that have replicated within the cell. We have cataloged the CRISPR spacer sequences from cellular metagenomic data from high-temperature (>80°C), acidic (pH < 4) hot spring environments located in Yellowstone National Park (YNP). We designed a microarray platform utilizing these CRISPR spacer sequences as potential probes to detect viruses present in YNP hot spring environments. We show that this microarray approach can detect viral sequences directly from virus-enriched environmental samples, detecting new viruses which have not been previously characterized. We further demonstrated that this microarray approach can be used to examine temporal changes in viral populations within the environment. Our results demonstrate that CRISPR spacer sequence-based microarrays will be useful tools for detecting and monitoring viruses from diverse environmental samples.

Viruses are arguably the most abundant lifelike entities on the planet (2, 29, 33); however, there are few effective approaches that characterize the diversity and composition of viruses in any particular environment. Using our environmental data, we estimate that only 0.01 to 0.1% of all viruses present within acidic hot springs have been isolated to date (unpublished data). One reason for our poor understanding of virus diversity is the difficulty of isolating and maintaining potential cellular hosts in culture. This is especially the case for microbial hosts from extreme environments. In an effort to overcome some of the inherent limitations of culture-dependent approaches, culture-independent viral metagenomic approaches have been undertaken to access viral diversity within selected environments (1–5, 7–16, 18, 22, 28, 29, 33, 35). These studies have generally found very high levels of viral diversity. While a highly useful approach, viral metagenomic studies are both time-consuming and expensive to perform.

The CRISPR (clusters of regularly interspaced short palindromic repeats)/Cas (CRISPR-associated proteins) system is a broadly distributed microbial immunity system to defend against invading nucleic acids (viruses, plasmids, and mobile genetic elements) and is thought to operate in many bacteria and most archaeal species (6, 14, 17, 19–21, 23, 32, 34). The CRISPR/Cas system has been found in 40% of sequenced bacterial genomes and 90% of sequenced archaea from genome bioinformatic analysis (20). Our mechanistic understanding of the CRISPR/Cas system is rapidly emerging, but it is currently incomplete (6, 17). The current model for the functioning of the CRISPR/Cas system has three basic components, which are (i) the recognition of the invading nucleic acid (the viral genome) and the incorporation of short (32- to 36-bp) nucleotide sequences derived from viral genome sequences (termed spacer sequences) separated by approximately 45 nucleotides (nt) of host-derived direct repeat sequences within the cellular CRISPR loci, (ii) the transcription of the CRISPR loci to produce small (68- to 75-bp) RNA fragments that span from one direct repeat unit to the adjacent repeat unit with the spacer sequence in between, and (iii) the incorporation of the small RNA sequences into Cas-derived protein complexes that are capable of targeting and inactivating viral genomes that correspond to the viral spacer sequence. The number of CRISPR loci per genome varies from organism to organism, as does the total number of CRISPR repeat units/spacer units per locus. Some CRISPR loci are very short (a single repeat unit and spacer sequence), while others are long arrays of hundreds of repeat units and spacer sequences. The repeat units within each CRISPR locus are almost always identical to each other, but the spacer sequences are unique. From the 59 archaeal and 602 bacterial sequenced genomes containing CRISPR, a catalog of more than 28,500 unique spacer sequences can be created (http://crispr.u-psud.fr/crispr/). In addition, metagenomic studies from low-complexity microbial community environments such as acid mine drainage (1) or acidic hot springs (A. C. Ortmann, M. M. Bateson, B. Bolduc, A. Mazurie, V. J. B. Ruigrok, F. F. Roberto and M. J. Young, unpublished data) can yield upward of 10,000 unique CRISPR loci spacer sequences. The vast major-

* Corresponding author. Mailing address: 119 Plant Biosciences Building, Montana State University, Bozeman, MT 59717. Phone: (406) 994-5158. Fax: (406) 994-1975. E-mail: myoung@montana.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 17 September 2010.

ity of the CRISPR spacer sequences do not correspond to either known cellular or viral sequences present in the public databases (30), suggesting that most CRISPR spacer sequences are derived from viral genomes which have not yet been described. The CRISPR spacer sequences are unique within an individual bacterial or archaeal genome and may even be unique within individual cells, with the possibility that each cell within the population has its own unique CRISPR loci comprised of a unique set of spacer sequences. Recent studies have also suggested that the CRISPR loci can be quite dynamic, both incorporating new and removing previously existing spacer sequences quickly (1, 14, 30). Taken together, this suggests the possibility that each CRISPR spacer sequence corresponds to a unique viral genome and may be used as a probe to detect previously unknown viruses.

We examined whether CRISPR spacer sequences could be used to detect unknown archaeal viruses in an extreme thermal environment. In comparison to the viruses of *Eukarya* and *Bacteria*, very little is known about the viruses that infect *Archaea*. Of the estimated 5,100 known viruses, only 37 have been isolated from *Archaea* (http://www.ncbi.nlm.nih.gov/ICTVdb/). In spite of our inadequate knowledge of these viruses, their discovery and characterization have significantly increased our understanding of viral ecology and evolution. The high-temperature (>80°C), acidic (pH < 4.0) hot springs found within Yellowstone National Park (YNP) are dominated by *Archaea* and their viruses, and as such these hot springs provide ideal environments for searching for new archaeal viruses. The goals of this study were to determine if a microarray approach based on cellular CRISPR spacer sequences as probes can be used to detect unknown viruses directly from YNP hot spring environmental samples and as a new tool to monitor temporal changes in virus populations.

MATERIALS AND METHODS

Hot spring study sites and sources of CRISPR spacer sequences. Two high-temperature, acidic hot springs were used in this study. The first is an acidic (~pH 2.5) hot spring (~82°C; designation CHANN041; 44.6532°N, 110.4847°W) located in the Crater Hills area of Yellowstone National Park. Five samples collected from CHANN041 at different times were used as a source of environmental viral DNA for hybridization to CRISPR spacer sequence-based microarray platforms (described below). In previous studies, CHANN041 was used for generating both cellular and virus-enriched environmental metagenomic data sets (see Table S1 in the supplemental material) (Ortmann et al., unpublished). CHANN041 is a vapor-dominated acid-sulfate hot spring approximately 3 m by 21 m in size that does not have major inflow or outflow channels and that is continuously mixing due to excessive outgassing. A cellular environmental metagenomic data set from a second hot spring, Nymph Lake 10 (NL10; ~91°C; pH 3.5 to 4.5; 44.7536°N, 110.7237°W), was used as a source for additional cellular CRISPR spacer sequences (see Table S1 in the supplemental material). The individual reads from the cellular environmental metagenomic data sets from CHANN041 and NL10 were used to identify cellular CRISPR-related sequences that were used to create a library of CRISPR spacer sequences using the CRISPR finder on the CRISPR home page (http://crispr.u-psud.fr/crispr/).

Virus-enriched environmental samples. Environmental samples collected from CHANN041 were enriched for viruses. Briefly, cellular and viral fractions were separated by size filtration (Fig. 1). The virus-enriched fraction was the result of filtering hot spring water through two successive 0.8- and 0.2-μm filters (Pall, Port Washington, NY) and retaining the flowthrough in sterile 26-ml ultracentrifuge tubes. Upon return to the laboratory, the virus-enriched filtrate was centrifuged at 25,000 rpm for 2 h to pellet virus particles. The virus pellets were serially resuspended in 250 μl double-distilled water (ddH_2O) (Sigma, St. Louis, MO) and stored at −80°C until needed. Virus-enriched samples from CHANN041 were collected on 1 August 2008, 13 October 2008, 14 January 2009, 9 September 2009, and 18 February 2010.

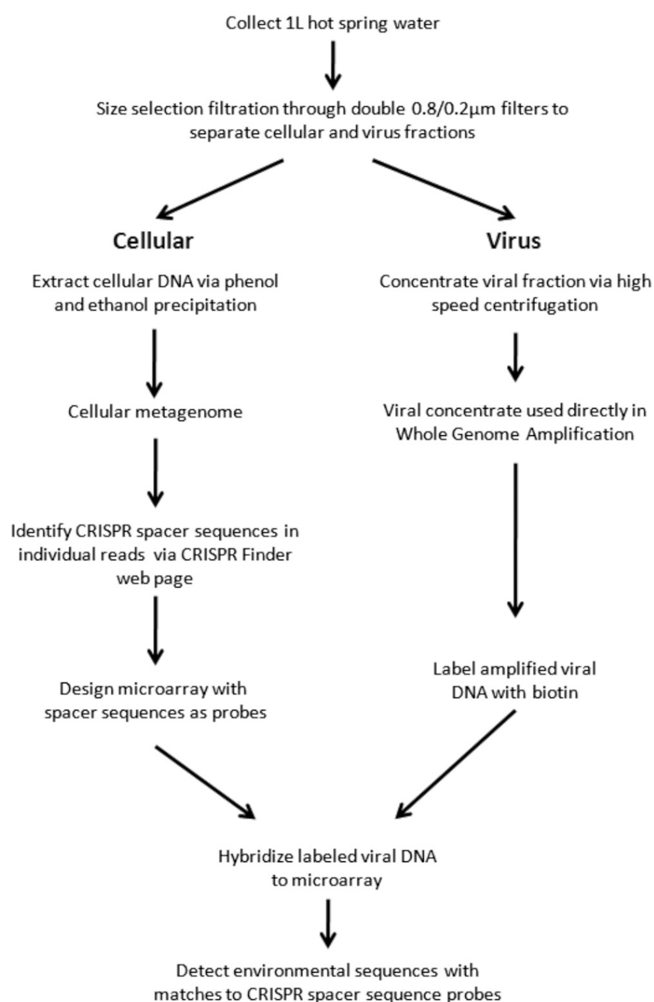

FIG. 1. Overall schematic representing the methods used in this study.

Microarray design. Two microarray designs were utilized. For the first-generation CRISPR spacer-based microarray, the platform was the CombiMatrix (Mukilteo, WA) 4X2K custom array. This microarray design is composed of four separate arrays, each with 2,240 spots. For the initial round of experiments, the four arrays each contained identical sequences, allowing one to process four independent samples on the same microarray. The following four groups of 35-bp oligonucleotide sequences were placed on the array: (i) sequences corresponding to controls of archaeal 16S ribosomal DNA (rDNA) sequences derived from YNP metagenomic data sets; (ii) sequences corresponding to the *Arabidopsis thaliana* plant genome; (iii) known *Sulfolobus* viral sequences (*Sulfolobus* spindle-shaped virus [SSV1] [25], *Sulfolobus islandicus* rod-shaped virus [SIRV1] [26], and *Sulfolobus* turreted icosahedral virus [STIV] [27]); and (iv) CRISPR spacer sequences identified in environmental metagenomic data sets from CHANN041 (18; Ortmann et al., unpublished) and Nymph Lake (unpublished data) (Table 1). Each CRISPR spacer sequence is represented in four sequence forms on the array: perfect match (PM), 1-nt mismatch (MM), antisense perfect match (anti-PM), and 1-nt antisense mismatch (anti-MM). The cellular spacer sequences from CHANN041 that perfectly matched the paired viral metagenomes (Ortmann et al., unpublished) by BLAST analysis were preferentially chosen for the microarray. All oligonucleotides placed on the array were designed to have similar base compositions and predicted melting temperatures (~65°C). The second generation of microarrays was expanded to include all of the CRISPR spacer sequences derived from CHANN041 and NL10 cellular metagenomes on the CombiMatrix 4X2K custom array. With this array design, only two samples could be hybridized on each microarray, and each sample was inoculated onto two separate arrays. For the second-generation microarrays,

TABLE 1. Probe category and the number of probes on the two generations of microarrays

Probe category	No. of probes	
	First generation[a]	Second generation[b]
16S rDNA sequences	104	120
CHANN041 CRISPR spacer sequences	1,855[c]	3,706
NL10 CRISPR spacer sequences	154[d]	528
SIRV1	50	50
SSV1	33	32
STIV	36	36
Arabidopsis	8	8
Total	2,240	4,480

[a] The first-generation arrays contain 4 probes per CRISPR spacer sequence: perfect match, mismatch, anti-perfect match, and anti-mismatch.

[b] The second-generation arrays contain 2 probes per CRISPR spacer sequence: perfect match and mismatch.

[c] There are 459 CHsp that have 4 probes per CRISPR spacer sequence; 19 CHsp were repeated to fill in spots on the array.

[d] There are 38 NLsp that have 4 probes per CRISPR spacer sequence; 2 NLsp were repeated to fill in spots on the array.

35-nt CRISPR spacer sequences were represented as only perfect match and 1-bp mismatch probes. The 16S rDNA sequences, *Arabidopsis* sequences, and *Sulfolobus* virus sequences were also included on the microarrays as controls (Table 1). Sequences of all probes used on both generations of arrays are provided in the probe list files in the supplemental material (array1_probelist.txt and array2_probelist.txt).

Virus-infected *Sulfolobus* cultures. Cultures of *Sulfolobus solfataricus* strain P2[3] (24) either infected with STIV or noninfected were used as controls to test the validity of the microarray. *S. solfataricus* P2[3] was cultured in medium 182 (http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium182) at 85°C and a pH of 2.5. When the culture reached mid-exponential phase (optical density at 650 nm [OD_{650}] of ~0.2), it was infected with STIV at a multiplicity of infection (MOI) of ~1.7. Forty-eight hours postinfection, the culture was processed in the same manner as environmental samples to separate cellular and viral fractions (described above).

Preparation of samples for hybridization to CRISPR spacer sequence microarrays. The preparation of biotin-labeled viral DNA for hybridization to CRISPR spacer sequence microarrays was performed as recommended by CombiMatrix. Briefly, 10 μl of virus-enriched sample was used in a whole-genome amplification (WGA) reaction according to manufacturer protocols (Sigma). After amplification, small oligonucleotides and nucleotides were removed using the QIAquick PCR purification kit (Qiagen, Valencia, CA). The amplified DNA was used in a biotin-labeling reaction according to manufacturer protocols (Mirius, Madison, WI). The labeling reaction was subsequently treated to remove unincorporated biotin using the QIAquick nucleotide removal kit (Qiagen) and eluted in a final volume of 30 μl ddH_2O. Replicates of individual samples were hybridized with similar concentrations of labeled DNA.

The prehybridization and hybridization steps were performed as recommended by the manufacturer (CombiMatrix). Briefly, the microarray was rehydrated prior to hybridization by incubation in nuclease-free water at 65°C for 10 min. The hybridization chambers were filled with 30 μl of prehybridization solution (see the supplemental material for all solution components) and incubated with gentle rotation at 50°C for 30 min. The hybridization chambers were then filled with hybridization solution and incubated overnight at 50°C. After hybridization for 14 to 16 h, the posthybridization washing and posthybridization labeling with FluoroLink Cy5-labeled streptavidin (GE Healthcare/Amersham Biosciences, Piscataway, NJ) were carried out as directed by CombiMatrix protocols. Briefly, the hybridization chambers were sequentially washed with 6× SSPE-Tween 20 (SSPET; 1× SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA [pH 7.7]), 3× SSPET, 0.5× SSPET, and phosphate-buffered saline–Tween 20 (PBST) for 1 min each at room temperature (see the supplemental material for the compositions of washes). After the chambers were washed, biotin-blocking solution was added to the hybridization chambers and incubated for 15 min at room temperature. The dye-labeling solution was added to the chambers, and the array was incubated in the dark at room temperature for 30 min (see the supplemental material for the compositions of solutions). The hybridization chambers were then washed with PBST and kept in PBS until imaged. Microarray chips were reused up to four times by following the stripping protocols of the hybridized DNA supplied by the manufacturer (CombiMatrix; see the supplemental material).

Imaging and data analysis. Microarrays were imaged following recommended protocols from the manufacturer (CombiMatrix). Briefly, a few drops of imaging solution (provided by CombiMatrix) were added to the arrays, and the arrays were placed in a GenePix 4000B scanner (Molecular Devices Corp., Sunnyvale, CA), available at the Functional Genomics Core Facility at Montana State University (Bozeman, MT). The instrument settings were as follows: photomultiplicator (PMT) gain, 650; laser power, 100%; pixel size, 5; focus position, 100; and lines to average, 2. The image was saved as a single-image TIF file and then analyzed in GenePix Pro version 6.1 (Molecular Devices Corp.). The array design files (supplied by CombiMatrix) were converted to a GenPix array list (GAL) file using the GAL file tool (available from CustomArray, Inc.). The data were extracted and analyzed using Excel 2007 (Microsoft). The mean background fluorescence was subtracted from the mean fluorescence for each spot on the array. The spots that resulted in significant fluorescence above the background were used in further analyses.

RESULTS

Microarray design. The first-generation CRISPR spacer sequence microarray was designed to test the ability of a microarray approach to detect viruses in control and environmental samples. The first-generation microarray consisted of a total of 2,240 oligonucleotide probes. From a total of approximately 2,000 unique spacers, 460 CRISPR spacer sequences were selected from CHANN041 cellular environmental metagenomic data sets (CH2007_04_cellular and CH2007_09_cellular), which resulted in 1,855 probes (4 probes per CRISPR spacer sequence [PM, MM, anti-PM, and anti-MM probes]; 19 spacers were used twice in order to fill spots on the array). There were only 38 CRISPR spacer sequences collected from the NL10 hot spring cellular metagenomic data set (NL2007_08_cellular), resulting in a total of 154 Nymph Lake-specific probes (PM, MM, anti-PM, and anti-MM probes; two NL spacers were repeated). Controls on each array included 104 probes to 4 of the 16S rDNA genes identified in the CHANN041 and NL10 hot spring cellular metagenomes, 119 probes to known *Sulfolobus* viral genomes (50 SIRV1, 33 SSV1, and 36 STIV probes) that resulted in, on average, one probe per viral genome open reading frame (ORF), and 8 probes to the *Arabidopsis* plant genome (Table 1).

The second-generation microarray was designed to detect a broader diversity of unknown viruses and to follow temporal changes in viral population within the CHANN041 hot spring. The second-generation arrays contained a total of 4,480 oligonucleotide probes. This included 1,853 CRISPR spacer sequences selected from the CHANN041 cellular environmental metagenomic data set, which resulted in 3,706 probes (2 probes per CRISPR spacer sequence [PM and MM]; one CHANN041 CRISPR spacer sequence [CHsp] was repeated). There were 264 CRISPR spacer sequences selected from the NL10 hot spring cellular metagenomic data set and newly created PCR libraries (unpublished data), resulting in a total of 528 NL10-specific probes. The controls included on each second-generation array were in the same categories as those used on the first-generation arrays, but the numbers of included sequences differed (Table 1).

Proof-of-concept experiments using the first-generation microarray platform. Three sets of controls were used to test the validity of a cellular CRISPR spacer sequence-based microarray to detect viruses. The first set of control experiments compared the virus-enriched fractions of *Sulfolobus* laboratory cul-

TABLE 2. Number and percentage of positive sequences detected in proof-of-concept experiments using the first-generation arrays

Probe[a]	No. (%) of positive sequences from[b]:			
	SsoP2[3] plus STIV	SsoP2[3] without STIV	NL plus STIV	NTC
16S rDNA	2 (2)	3 (3)	8 (8)	0 (0)
CHsp	76 (17)	2 (0.4)	23 (5)	2 (0.4)
NLsp	9 (23)	0 (0)	4 (11)	1 (3)
STIV	23 (64)	0 (0)	15 (42)	0 (0)
SSV1	0 (0)	0 (0)	1 (3)	0 (0)
SIRV1	0 (0)	0 (0)	2 (4)	0 (0)

[a] CHsp, spacers from the Crater Hills cellular metagenomes (collected in April and September 2007); NLsp, spacers from the Nymph Lake cellular metagenome (collected in August 2007).

[b] Numbers in parentheses are the percentages of probes detected from the total number in each category. SsoP2[3] plus STIV[c] and SsoP2[3] without STIV, *S. solfataricus* cultures infected and noninfected with STIV, respectively; NL plus STIV, hot spring water collected from Nymph Lake spiked with STIV particles; NTC, no-template control (WGA amplification of sterile water).

tures that were infected with STIV to *Sulfolobus* cultures that were not infected with STIV. The purpose of these experiments was to establish that our protocols for separating cellular and viral DNA were sufficient to specifically detect viral sequences (STIV) with a relatively low background of cellular sequence detection. The laboratory cultures were processed in a manner identical to that for environmental hot spring samples. No viral sequences were detected by the microarray when using DNA from the *Sulfolobus* cultures that were not infected with STIV particles (Table 2). Sixty-four percent of the probes (23/36) directed against the STIV genome were detected in *Sulfolobus* cultures that had been infected with the virus (Table 2). Not surprisingly, it was not possible to eliminate the detection of all cellular signals present within the viral fraction. A small number of the 16S rDNA probes were positive (2 to 3%) (Table 2). As expected, no probes to *Arabidopsis* sequences were detected. These results indicate that there was little nonspecific or cellular DNA hybridization detected by the CRISPR spacer sequence microarrays.

The second proof-of-concept experiment involved spiking hot spring water with a known amount of virus. STIV purposely added directly to hot spring environmental samples was detectable by the CRISPR spacer sequence array (Table 2). An environmental sample of NL10 hot spring water (500 ml) was spiked with 4.4×10^8 particles of STIV, and 500 ml was filtered to create a virus-enriched fraction. DNA from the virus-enriched fraction was prepared for application on the microarray platform. Labeled DNA (21 ng) was used in the hybridization reaction, which we estimate represents DNA from 1.5×10^8 virus particles, assuming 100% recovery. We were able to detect 15 STIV sequences out of the 36 probes present on the microarray (Table 2).

The third set of controls tested for the effect of nonspecific amplification as a result of using a WGA step in the DNA-labeling protocol. Only a low level of nonspecific background was evident as a result of the WGA reaction. Control reactions of substituting sterile water for the virus-enriched template were performed in two independent WGA reactions. After the WGA reaction, the total amount of synthesized DNA was 310 ng. The maximum volume was added to the biotin-labeling reaction, which represented approximately 80 ng and after purification resulted in a final yield of 39 ng of biotin-labeled DNA. The maximum volume of biotin-labeled DNA was subsequently used to hybridize to the microarray, which resulted in 7.8 ng of the no-template control (NTC) material being hybridized to the microarray. Of the 2,240 spots on the microarray, only three spots were above background levels when the NTC material was applied (Table 2).

Detection of spacer sequences from environmental samples. A schematic overview of the experimental design to detect viral sequences from environmental samples using the CRISPR spacer sequence microarray platform is provided (Fig. 1). The detection of potentially unknown viruses in CHANN041 was examined using the first-generation CRISPR spacer sequence microarray platform. The CRISPR spacer sequence microarray was used to detect viral sequences within the virus-enriched fraction from CHANN041 collected on 1 August 2008. In triplicate, 10 μl of the virus-enriched fraction was used in independent WGA reaction mixtures (the triplicates will be referred to as CH0808A to CH0808C). The resulting DNA yields after WGA ranged from 730 to 765 ng. Approximately 60 ng of biotin-labeled material was subsequently hybridized to three individual CRISPR spacer sequence microarrays at 50°C for 14 h. The total number of hybridized probes above background levels ranged from 491 to 558 (data not shown). Of those spots, 236 to 261 represented unique spacer sequences, with the remainder being redundant PM or anti-PM probes that were also positive (see Table S2 in the supplemental material). Single-base-pair-mismatched probes were often detected (data not shown). A total of 221 unique sequences were common among the three independent replicates. Five to 17 probes appeared in only one of the three independent replications, indicating consistency between the triplicate reactions (data not shown). Of the 221 unique sequences shared in the three independent replicates, 195 were detected by probes derived from CHANN041 CRISPR spacer sequences, 8 sequences were detected by probes derived from NL10 spacer sequences (NLsp), 5 by STIV probes, 6 by SIRV1 probes, 3 by SSV1 probes, and 4 by 16S rDNA probes (see Table S2 in the supplemental material). Overall, this represents ~42% of the total unique CHsp present on the microarray.

There was a significant overlap between the viral sequences detected by the CRISPR spacer sequence microarray and the viral population assessed by viral metagenomics (Ortmann et al., unpublished). On average, 61% of the viral sequences detected by the CRISPR spacer sequence microarray matched viral sequences independently determined in CHANN041 by using a viral metagenomics approach (Table 3), indicating agreement between the viral metagenomics-based approach and the CRISPR spacer sequence microarray approach. This high degree of overlap is even more remarkable given that there was at least 11 months separating the sampling dates of the hot spring used for this microarray analysis and the viral metagenomic analysis.

The CRISPR spacer sequence-based microarray was used to detect viral sequences in a second CHANN041 virus-enriched sample collected 5 months later, on 14 January 2009. In triplicate, 10 μl of the virus preparation was used in the WGA reaction (the triplicates will be referred to as CH0901A to CH0901C). The resulting DNA yields were 5.8 to 6.8 μg. One microgram of the amplified DNA was used in a biotin-labeling

TABLE 3. Correspondence between first-generation CRISPR spacer sequence-based microarrays and viral metagenomics approaches

Fraction	No. of CH CRISPR spacer sequences[c]	No. (%) of CHsp that match viral metagenome[d]
CH0808[a]	217 ± 10	132 ± 5 (61)
CH0901[b]	207 ± 7	128 ± 2 (60)

[a] Virus-enriched fraction from CHANN041 sampled in August 2008.
[b] Virus-enriched fraction from CHANN041 sampled in January 2009.
[c] The number of CRISPR spacer sequences placed on the microarray that correspond to viral sequences present in the CHANN041 viral metagenome produced from virus samples obtained in April and September 2007. Values are reported as the averages and standard deviations of results from triplicate reactions.
[d] The number of positive CRISPR spacer sequence probes that corresponded to virus sequences present in the viral metagenome. Values are reported as the averages and standard deviations of results from triplicate reactions.

reaction for hybridization onto the CRISPR spacer sequence microarray. Biotin-labeled material (50 to 150 ng) was hybridized to three individual arrays at 50°C for 14 h.

There were 206 unique sequences detected across the three independent microarray assays (see Table S3 in the supplemental material). The total number of positive probes was 521, with many of these representing redundant probes (PM, MM, etc.) to CRISPR spacer sequences (data not shown). There were very few positive probes that appeared in only one of the three triplicate experiments (2–6). Of the 206 common sequences detected on independent arrays, 187 were detected by CHsp probes, 9 by NLsp probes, 4 by STIV probes, 3 by SSV1 probes, 1 by a SIRV1 probe, and 2 by 16S rDNA probes (see Table S3 in the supplemental material). On average, 60% of the CHsp-positive probes matched viral sequences present in the viral metagenome (Table 3), indicating agreement between the viral metagenome and the CRISPR spacer sequence microarray approach. Overall, these results indicate that the CRISPR spacer sequence microarray can reproducibly detect viral sequences in an environmental sample.

Detection and temporal analysis of viruses present in the CHANN041 population. In order to determine whether the CRISPR spacer sequence microarray can be used to monitor changes in the CHANN041 virus population over time, the two virus-enriched samples discussed above, which were collected 5 months apart, were compared (see Table S4 in the supplemental material). Of the approximately 424 total Crater Hills CRISPR spacer sequences that were detected at the two time points, 133 (35%) were detected at both time points (see Table S4 in the supplemental material). Of the 133 shared CRISPR spacer sequences between the samples collected at the two time points, 113 matched the viral sequences present in viral metagenomes independently produced from April and September 2007 sampling time points of CHANN041. In contrast, there were 59 CRISPR spacer sequence-positive probes that were unique to the August 2008 virus-enriched sample and 54 CRISPR spacer sequence-positive probes that were detected only at the January 2009 time point (see Table S4 in the supplemental material). These results indicate that the CRISPR spacer sequence microarray can detect temporal changes in the viral populations present in the environment.

The detection and temporal dynamics of hot spring viral assemblages in CHANN041 were further examined using the second-generation CRISPR spacer sequence microarray platform, which included 4 times as many CRISPR spacer sequences (Table 1). Four CHANN041 sampling time points that spanned 17 months from October 2008 to February 2010 were examined. For each time point, equal amounts of biotin-labeled viral DNA (1 μg) were hybridized to separate CRISPR spacer sequence microarrays. Independent replicate assays were carried out for each time point. The number of viral sequences detected by the expanded CRISPR spacer sequence microarray varied between each sampling time point. This variation ranged from 244 positive CHANN041 CRISPR spacer sequence probes for the October 2008 sampling date, 650 positive probes for the January 2009 sampling date, 525 positive probes for the September 2009 sampling date, and 437 positive probes for the February 2010 sampling date (Fig. 2; see also Tables S5 to S8 in the supplemental material). Across the four time points, 186 (~10%) of the total CHANN041 spacers represented on the microarray were common to all sampling dates (Fig. 2; see also Table S9 in the supplemental material). Surprisingly, there does not always appear to be a correlation between the proximity of sampling time points and the overlap in shared viral sequences detected. For example, there appears to be a time correlation of shared viral sequences detected between the 6 months separating the September 2009 and February 2010 sampling time points (359 shared viral sequences) compared to the 17 months separating the October 2008 and February 2010 sampling time points (186 shared viral sequences) (Fig. 2). However, there is a lack of correlation of shared viral sequences between the 9 months separating January 2009 and September 2009 (437 shared sequences) com-

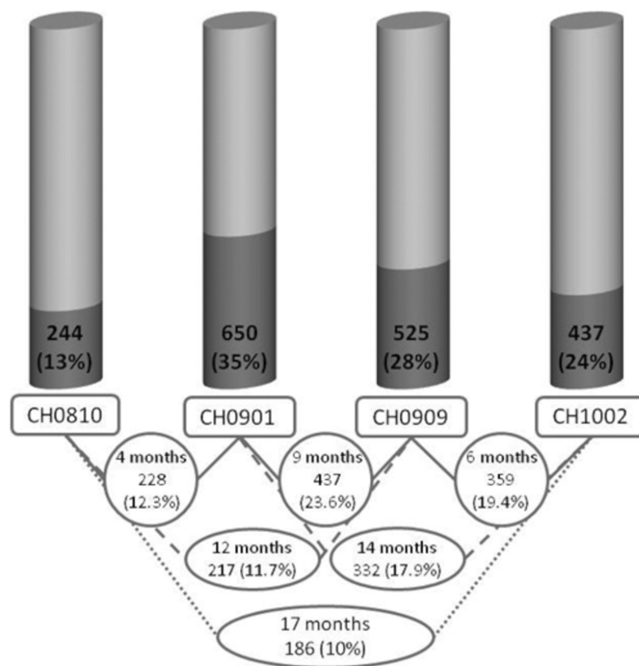

FIG. 2. Representation of the change in CHANN041 environmental sequences detected over time using the CRISPR spacer microarray. The percentages represent the proportion of the total CHANN041 spacers represented on the array (1,853) detected at any given time.

pared to the 4 months separating October 2008 and January 2009 (228 shared sequences) (Fig. 2). Of the 186 total detected positive CHANN041 CRISPR spacer sequence probes over all sampling times, 64 (34%) spacers matched viral sequences detected within the viral metagenomes produced from the April and September 2007 sampling time points (see Table S9 in the supplemental material).

Detection of CHANN041 viruses using NL10 CRISPR spacer sequences. It is likely that CHANN041 and NL10 hot springs have some overlap in their respective viral communities. One of the NL10 spacers matches a viral contig sequence from a viral matagenome collected from CHANN041 in January 2009 at 100% nucleic acid identity. We have found that several CRISPR spacer sequences from NL10 detect viral sequences in the virus-enriched fractions from CHANN041. In the first-generation CRISPR spacer sequence microarray, there were 38 unique (154 total spacer sequences, including PM, MM, anti-PM, anti-MM) spacer sequences derived from NL10. A total of 17 of these NL10 spacer sequences hybridized to CHANN041 virus-enriched samples (see Tables S2 and S3 in the supplemental material). Between the two CHANN041 viral samples tested, there was only one common NL10 spacer sequence detected over time (data not shown).

In the second-generation CRISPR spacer sequence microarray, there were 264 unique spacer sequences derived from the NL10 hot spring (528 total spacer probes, including PM and MM probes). A total of 215 of these spacer sequences were detected in the CHANN041 virus-enriched samples over time (see Tables S5 to S8 in the supplemental material). However, only 17 of the detected NL10 spacers were common in samples collected at the four CHANN041 time points that were hybridized to the second-generation array (data not shown).

DISCUSSION

Currently, there are few effective approaches to detect unknown viruses from environmental samples even though they are likely the most abundant lifelike entities present in most environmental samples (2, 33). The extreme environments that *Archaea* often inhabit make discovering new viruses even more challenging because of the difficulty of culturing organisms from these environments. Total viral metagenomic approaches provide one avenue for detecting new viral genotypes, but this approach still suffers from being relatively expensive in terms of both time and sequencing costs. The recent discovery and mechanistic understanding of the cellular CRISPR/Cas system provide a complementary approach for the detection of unknown viruses from environmental samples. To that end, we demonstrate here that a CRISPR spacer sequence-based microarray platform can be used to detect unknown viral genes in a high-temperature, acidic hot spring environment and to monitor temporal changes in the viral community.

We have established that the CRISPR/Cas spacer sequence microarray platform is capable of detecting virus sequences from environmental samples. Initial laboratory culture-based experiments established that this microarray approach could detect a known virus in culture and a virus spiked into an environmental sample (Table 2). While the separation of cellular and virus-enriched fractions is not complete (e.g., cellular sequences are still present in the viral fractions), the amount of cellular sequence carryover is sufficiently low to allow the detection of viruses. It is unlikely that there is a significant amount of the cellular CRISPR spacer sequences detected by the microarray assay since only a subset of the probes were positive at a given sampling time and they were significantly different in positive probes between time points. There is high reproducibility between independent replicates, and relatively small volumes of environmental sample are required.

The CRISPR spacer sequence microarray is useful for examining dynamics in viral community populations. Over the course of 17 months (October 2008 to February 2010), viral sequences from CHANN041 were detected by only 186 CRISPR spacer sequences across all time points, indicating a dynamic viral population structure (Fig. 2; see also Tables S9 and S10 in the supplemental material). The comparison of these time points with the CRISPR spacer sequences collected from the host metagenomes (collected in April and September 2007) reveals that 64 of the 186 shared CHANN041 CRISPR spacer sequence probes match the viral metagenome (~34%). This shared group of CRISPR spacer sequence probes may represent a viral population present in the hot spring continuously, and therefore the host population is under a selection pressure to maintain these CRISPR spacer sequences. However, the number of spacers detected varies at only one of the sampling times (see Table S9 in the supplemental material); these sequences may represent more-transient genes. The CRISPR spacer sequences detecting these viruses are not necessarily under strong selection pressure to be maintained in the cellular population. We speculate that these transient viruses are not present in the hot spring for continuous periods of time and are therefore difficult to detect using traditional methods.

Previous cellular metagenomic studies allowed us to create a library of approximately 2,000 CRISPR spacer sequences derived from YNP hot spring environments (18; Ortmann et al., unpublished). We suspect that multiple cellular CRISPR spacer sequences are present in different regions of the same viral genome, likely reducing the complexity of the true viral population recorded within the cellular CRISPR loci. On the other hand, our microarray hybridization conditions do not require that there be a 100% match between probe and target viral DNA in order for viral genome variants to be detected, allowing for the detection of viral genome variants within the environmental sample that are likely to exist. For the purpose of the proof-of-concept experiments, we intentionally biased the selection of CRISPR spacer sequences included on the first-generation microarray to ones that were known to match sequences present in our viral metagenomic data sets (collected in April and September 2007). As expected, we did find a strong correlation between viral sequences detected by using the CRISPR spacer sequence microarray approach and those detected by using the viral metagenomic approach from the same hot spring environment (Table 3). The percentage of CRISPR spacer sequence probes that matched the viral metagenomes was higher when the first-generation arrays were used than when the second-generation microarray platform was used (~60% versus ~30%), because the preselection of CRISPR spacer sequence probes on the first-generation microarray was biased toward those that had 100% matches to the viral sequences detected in viral metagenomic studies. In the second-generation microarray design, all usable CRISPR

spacer sequences were placed on the array, thereby eliminating the bias toward spacers that match the viral metagenomes.

It is evident that the majority of the CRISPR spacer sequence probes did not detect their corresponding target in the virus-enriched fractions tested. The overall average percentage of CRISPR spacer sequence probes that detected viral sequences using the second-generation CRISPR spacer sequence arrays was 28%. This suggests several possibilities. The first possibility is that the CRISPR spacer sequences present in the cellular metagenomes are to viruses that were no longer present or dominant in the CHANN041 hot spring at the time of sampling. This is what one might expect to observe if the viral population structure is quite dynamic. In support of this possibility, it is important to note that the cellular metagenomic sampling dates (April and September 2007 for CHANN041 and August 2007 for NL10) were well separated from the dates of sampling for the virus-enriched fractions using both CRISPR spacer sequence microarray platforms (CHANN041 sampling dates August 2008 through February 2010). Likewise, the differences in the specific positive probes detected between the four samples using the second-generation CRISPR microarray approach also suggest a dynamic viral population structure. Future higher-resolution temporal sampling using the CRISPR spacer sequence microarray should be insightful in further understanding the dynamics of these viral populations within their natural environments. Second, the CRISPR spacer sequences selected to be on the array could be to nonviral targets that are not present in the virus-enriched fractions. These nonrepresented targets could include other invading nucleic acids, such as plasmids and transposable elements that are also thought to be acted upon by the cellular CRISPR/Cas system. Third, the sensitivity of the CRISPR spacer sequence microarray approach may not be sufficiently high enough to detect low-abundance viruses, and/or the target DNA amplification and labeling strategy could be biased toward certain types of viral sequences present in the enriched viral fraction. Finally, the viral population may have diverged sufficiently between sampling points such that the selected CRISPR spacer sequence probes no longer detect their viral DNA. It is reasonable to assume that negative selection will act on members within the viral population that have sequences identical to active cellular CRISPR spacer sequence loci, resulting in a shift of the viral genome population structure that is able to avoid detection and possible inactivation by the cellular CRISPR/Cas system. A more expanded CRISPR spacer sequence microarray design, for example, one that has all possible CRISPR spacer sequences known or spacers from a closely related environment under study, should be quite useful in detecting unknown viruses and following viral population dynamics.

It is interesting to note that NL10-derived CRISPR spacer sequences detect virus from virus-enriched fractions collected from CHANN041. The two hot springs are separated by more than 30 km. Previous research has shown that virus movement between geographically isolated hot springs can occur and contributes to the viral diversity within YNP (31). Another possibility is that similar types of hot springs (those with similar pH values, temperatures, chemistries, etc.) have similar host populations which support overlapping virus populations. It is likely that our microarray-based approach is capable of detecting viruses that may be moving between hot springs within YNP.

Several studies have investigated temporal changes of virus populations within aquatic environments (7, 16, 28, 31). Most of these studies have shown a link in virus structure change to seasonality, time, or geographical location (7, 16, 28). In this study, we were able to design a microarray that is capable of detecting change in unknown virus populations within acidic hot springs. Further investigation is required to determine what environmental factors are driving the change in virus population. We believe that this microarray-based approach can be extended to other environments and other probe types as well. For example, instead of utilizing cellular spacer sequences, one could use viral metagenomic sequences collected from the environment of interest.

The CRISPR spacer sequence microarray has its limitations. One limitation of the CRISPR space sequence microarray assay is that it is not a quantitative measure of the amount of virus present. Second, the microarray platform detects only small segments (~35 nt) of viral genomes, and at present it is difficult to assign detection by multiple CRISPR spacer sequences to a common viral genome. It is envisioned that with future development, the CRISPR spacer sequence microarray assay will guide isolation of full-length viral genomes and eventually aid in the isolation of the viruses themselves. In addition, the CRISPR spacer sequence microarray requires prior knowledge of the repertoire of the cellular CRISPR sequences present in an environment, typically produced by investments in environmental metagenomic studies. However, the CRISPR spacer sequence microarray offers an attractive follow-up assay that is relatively inexpensive for detecting and monitoring unknown viruses in an environment. It is our goal to further develop the CRISPR spacer sequence microarray approach to aid in the discovery of novel viruses not only from extreme environments but from any environment of interest.

DNA microarrays have typically been used to detect differences in gene expression. However, Wang et al. designed a microarray for the detection of viral pathogens implicated in emerging infectious diseases (36). The advantage these researchers had in designing this array was the knowledge of the viral targets. Nevertheless, the success of the "virochip" illustrates that detecting viral sequences contained in a community sample using a DNA oligonucleotide microarray platform is possible. In addition to the virochip, other studies have illustrated the usefulness of a microarray-based platform for identifying viruses in a particular sample. For example, Breitbart et al. designed a microarray using viral metagenomic sequences collected from the gut of an infant (4). Using the microarray, they were able to show that the viral community within the gut of an infant dramatically changes in the first couple weeks of life (4). The results of our study extend the concept of using a microarray-based platform to detect previously unknown viruses in an environment and to monitor viral dynamics over time. It is envisioned that the further development of the CRISPR spacer sequence microarray will lead to more robust methods to isolate and characterize these unknown viruses. Future microarray designs include using longer oligonucleotide probes (~70 nt) designed from the viral metagenomes. With further development, we foresee using the

array as a "capturing device" by sequencing captured environmental viral sequences hybridized to the probes.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation grant numbers DEB-0936178 and EF-0802200 and the National Aeronautics and Space Administration grant number NNA-08CN85A. Montana State University's Functional Genomics Core Facility is supported by NCRR grant number P20 RR016455.

The research was conducted in Yellowstone National Park under the guidelines of permit number YELL-2010-SCI-5090.

REFERENCES

1. **Andersson, A. F., and J. F. Banfield.** 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science **320:** 1047–1050.
2. **Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle, and F. Rohwer.** 2006. The marine viromes of four oceanic regions. PLoS Biol. **4:**e368.
3. **Bench, S. R., T. E. Hanson, K. E. Williamson, D. Ghosh, M. Radosovich, K. Wang, and K. E. Wommack.** 2007. Metagenomic characterization of Chesapeake Bay virioplankton. Appl. Environ. Microbiol. **73:**7629–7641.
4. **Breitbart, M., M. Haynes, S. Kelley, F. Angly, R. A. Edwards, B. Felts, J. M. Mahaffy, J. Mueller, J. Nulton, S. Rayhawk, B. Rodriguez-Brito, P. Salamon, and F. Rohwer.** 2008. Viral diversity and dynamics in an infant gut. Res. Microbiol. **159:**367–373.
5. **Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer.** 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. **185:**6220–6223.
6. **Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. van der Oost.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960.
7. **Chen, F., K. Wang, S. Huang, H. Cai, M. Zhao, N. Jiao, and K. E. Wommack.** 2009. Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. Environ. Microbiol. **11:**2884–2892.
8. **Culley, A. I., A. S. Lang, and C. A. Suttle.** 2006. Metagenomic analysis of coastal RNA virus communities. Science **312:**1795–1798.
9. **Desnues, C., B. Rodriguez-Brito, S. Rayhawk, S. Kelley, T. Tran, M. Haynes, H. Liu, M. Furlan, L. Wegley, B. Chau, Y. Ruan, D. Hall, F. E. Angly, R. A. Edwards, L. Li, R. V. Thurber, R. P. Reid, J. Siefert, V. Souza, D. L. Valentine, B. K. Swan, M. Breitbart, and F. Rohwer.** 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature **452:**340–343.
10. **Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer.** 2008. Functional metagenomic profiling of nine biomes. Nature **452:**629–632.
11. **Dinsdale, E. A., O. Pantos, S. Smriga, R. A. Edwards, F. Angly, L. Wegley, M. Hatay, D. Hall, E. Brown, M. Haynes, L. Krause, E. Sala, S. A. Sandin, R. V. Thurber, B. L. Willis, F. Azam, N. Knowlton, and F. Rohwer.** 2008. Microbial ecology of four coral atolls in the Northern Line Islands. PLoS One **3:**e1584.
12. **Djikeng, A., R. Kuzmickas, N. G. Anderson, and D. J. Spiro.** 2009. Metagenomic analysis of RNA viruses in a freshwater lake. PLoS One **4:**e7264.
13. **Fierer, N., M. Breitbart, J. Nulton, P. Salamon, C. Lozupone, R. Jones, M. Robeson, R. A. Edwards, B. Felts, S. Rayhawk, R. Knight, F. Rohwer, and R. B. Jackson.** 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microbiol. **73:**7059–7066.
14. **Heidelberg, J. F., W. C. Nelson, T. Schoenfeld, and D. Bhaya.** 2009. Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. PLoS One **4:**e4169.
15. **Held, N. L., and R. J. Whitaker.** 2009. Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. Environ. Microbiol. **11:**457–466.
16. **Helton, R. R., and K. E. Wommack.** 2009. Seasonal dynamics and metagenomic characterization of estuarine viriobenthos assemblages by randomly amplified polymorphic DNA PCR. Appl. Environ. Microbiol. **75:**225–2265.
17. **Horvath, P., and R. Barrangou.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167–170.
18. **Inskeep, W. P., D. B. Rusch, Z. J. Jay, M. J. Herrgard, M. A. Kozubal, T. H. Richardson, R. E. Macur, N. Hamamura, R. Jennings, B. W. Fouke, A. L. Reysenbach, F. Roberto, M. Young, A. Schwartz, E. S. Boyd, J. H. Badger, E. J. Mathur, A. C. Ortmann, M. Bateson, G. Geesey, and M. Frazier.** 2010. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS One **5:**e9773.
19. **Karginov, F. V., and G. J. Hannon.** 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell **37:**7–19.
20. **Kunin, V., R. Sorek, and P. Hugenholtz.** 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. **8:**R61.
21. **Lillestol, R. K., S. A. Shah, K. Brugger, P. Redder, H. Phan, J. Christiansen, and R. A. Garrett.** 2009. CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol. Microbiol. **72:**259–272.
22. **Lopez-Bueno, A., J. Tamames, D. Velazquez, A. Moya, A. Quesada, and A. Alcami.** 2009. High diversity of the viral community from an Antarctic lake. Science **326:**858–861.
23. **Marraffini, L. A., and E. J. Sontheimer.** 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. **11:** 181–190.
24. **Ortmann, A. C., S. K. Brumfield, J. Walther, K. McInnerney, S. J. Brouns, H. J. van de Werken, B. Bothner, T. Douglas, J. van de Oost, and M. J. Young.** 2008. Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus* turreted icosahedral virus. J. Virol. **82:**4874–4883.
25. **Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W. D. Reiter, and W. Zillig.** 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. Virology **185:**242–250.
26. **Prangishvili, D., H. P. Arnold, D. Gotz, U. Ziese, I. Holz, J. K. Kristjansson, and W. Zillig.** 1999. A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. Genetics **152:**1387–1396.
27. **Rice, G., L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J. E. Johnson, T. Douglas, and M. Young.** 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc. Natl. Acad. Sci. U. S. A. **101:**7716–7720.
28. **Rodriguez-Brito, B., L. Li, L. Wegley, M. Furlan, F. Angly, M. Breitbart, J. Buchanan, C. Desnues, E. Dinsdale, R. Edwards, B. Felts, M. Haynes, H. Liu, D. Lipson, J. Mahaffy, A. B. Martin-Cuadrado, A. Mira, J. Nulton, L. Pasic, S. Rayhawk, J. Rodriguez-Mueller, F. Rodriguez-Valera, P. Salamon, S. Srinagesh, T. F. Thingstad, T. Tran, R. V. Thurber, D. Willner, M. Youle, and F. Rohwer.** 2010. Viral and microbial community dynamics in four aquatic environments. ISME J. **4:**739–781.
29. **Schoenfeld, T., M. Patterson, P. M. Richardson, K. E. Wommack, M. Young, and D. Mead.** 2008. Assembly of viral metagenomes from Yellowstone hot springs. Appl. Environ. Microbiol. **74:**4164–4174.
30. **Shah, S. A., N. R. Hansen, and R. A. Garrett.** 2009. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. Biochem. Soc. Trans. **37:** 23–28.
31. **Snyder, J. C., B. Wiedenheft, M. Lavin, F. F. Roberto, J. Spuhler, A. C. Ortmann, T. Douglas, and M. Young.** 2007. Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. U. S. A. **104:**19102–19107.
32. **Sorokin, V. A., M. S. Gelfand, and I. I. Artamonova.** 2010. Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. Appl. Environ. Microbiol. **76:**2136–2144.
33. **Suttle, C. A.** 2007. Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. **5:**801–812.
34. **van der Oost, J., M. M. Jore, E. R. Westra, M. Lundgren, and S. J. Brouns.** 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. **34:**401–407.
35. **Vega Thurber, R., D. Willner-Hall, B. Rodriguez-Mueller, C. Desnues, R. A. Edwards, F. Angly, E. Dinsdale, L. Kelly, and F. Rohwer.** 2009. Metagenomic analysis of stressed coral holobionts. Environ. Microbiol. **11:**2148–2163.
36. **Wang, D., L. Coscoy, M. Zylberberg, P. C. Avila, H. A. Boushey, D. Ganem, and J. L. DeRisi.** 2002. Microarray-based detection and genotyping of viral pathogens. Proc. Natl. Acad. Sci. U. S. A. **99:**15687–15692.