AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

VIRUS-CELL INTERACTIONS

An Uncultivated Virus Infecting a Nanoarchaeal Parasite in the Hot Springs of Yellowstone National Park

Jacob H. Munson-McGee,[a]* Colleen Rooney,[b]* Mark J. Young[c]

[a]Department of Microbiology and Immunology, Montana State University, Bozeman, Montana, USA
[b]Department of Chemistry and Biochemistry, Montana State University, Bozeman, Montana, USA
[c]Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, Montana, USA

ABSTRACT The Nanoarchaeota are small cells with reduced genomes that are found attached to and dependent on a second archaeal cell for their growth and replication. Initially found in marine hydrothermal environments and subsequently in terrestrial geothermal hot springs, the Nanoarchaeota species that have been described are obligate ectobionts, each with a different host species. However, no viruses had been described that infect the Nanoarchaeota. Here, we identify a virus infecting Nanoarchaeota by the use of a combination of viral metagenomic and bioinformatic approaches. This virus, tentatively named Nanoarchaeota Virus 1 (NAV1), consists of a 35.6-kb circular DNA genome coding for 52 proteins. We further demonstrate that this virus is broadly distributed among Yellowstone National Park hot springs. NAV1 is one of the first examples of a virus infecting a single-celled organism that is itself an ectobiont of another single-celled organism.

IMPORTANCE Here, we present evidence of the first virus found to infect Nanoarchaeota, a symbiotic archaean found in acidic hot springs of Yellowstone National Park, USA. Using culture-independent techniques, we provide the genome sequence and identify the archaeal host species of a novel virus, NAV1. NAV1 is the first example of a virus infecting an archaeal species that is itself an obligate symbiont and dependent on a second host organism for growth and cellular replication. On the basis of annotation of the NAV1 genome, we propose that this virus is the founding member of a new viral family, further demonstrating the remarkable genetic diversity of archaeal viruses.

KEYWORDS Nanoarchaeota, archaeal virus, virus of Nanoarchaeota., Nanoarchaea, virus of Nanoarchaea

Viruses are ubiquitous features of cellular life. Their discovery and characterization have often led to new insights into the biology of cells which they infect. Most of the viruses that infect parasites are bacteriophage that infect enodosymbiotic or disease-causing bacteria. Among the few systems where phage have been described that infect a single-celled organism that itself parasitizes another single-celled organisms is that of the *Bdellovibrio*, which are bacteria that prey upon other Gram-negative bacteria (1–3). In this system, the presence of the bacteriophage VL-1 reduces the titer of *Bdellovibrio*, causing more prey cells to survive and increasing the phage titer (2). However, the interactions between *Bdellovibrio* and their prey as well as between *Bdellovibrio* and their phage are more akin to classic predator-prey relationships than to a long-term parasitic relationship.

The Nanoarchaeota represent the first example of an archaeal species that is entirely reliant upon a second archaeal species for its growth and replication (4). Ever since their discovery in 2002, the Nanoarchaeota have sparked interest due to their highly reduced genomes, rapidly evolving gene sequences, symbiotic lifestyle, and phylogenetic place-

Citation Munson-McGee JH, Rooney C, Young MJ. 2020. An uncultivated virus infecting a nanoarchaeal parasite in the hot springs of Yellowstone National Park. J Virol 94:e01213-19. https://doi.org/10.1128/JVI.01213-19.

Editor Colin R. Parrish, Cornell University

Address correspondence to Mark J. Young, myoung@montana.edu.

* Present address: Jacob H. Munson-McGee, Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA; Colleen Rooney, Department of Computer Science and Computer Engineering, Portland State University, Portland, Oregon, USA.

Received 5 August 2019
Accepted 10 October 2019
Accepted manuscript posted online 30 October 2019
Published 17 January 2020

ment within the Archaea. Originally described as a deeply branching phylum of the Archaea, more recent analyses have classified the Nanoarchaeota as part of the DPANN archaeal superphylum. Despite rRNA surveys suggesting that Nanoarchaeota are widespread in marine and terrestrial environments, there are currently only three cultured representatives of the Nanoarchaeota. The original *Nanoarchaeum equitans* isolate was collected from a shallow marine hydrothermal vent (4) and the terrestrial species "*Candidatus* Nanopusillus acidilobi" from a circumneutral hot spring in Yellowstone National Park (YNP) (5), and the recently described species "*Candidatus* Nanoclepta minutus" and its host *Zestosphaera tikiterensis* were isolated from a hot spring in New Zealand (6). In addition, there are multiple examples of terrestrial Nanoarchaeota genomes that have been sequenced but that remain uncultured (7–9). A recent analysis of Nanoarchaeota and their cosorted putative archaeal hosts suggests a wide range of species within the Crenarchaeota with which Nanoarchaeota can associate. However, all Nanoarchaeota cells associated with a particular host cell type are clonal (7). The physical nature of the interaction between Nanoarchaeota cells and their host cells is under investigation. While membrane stretching between cells suggests a strong intercellular association that likely includes the sharing of metabolites and other small molecules, the extent of this association is not known (5).

While culture-based virus-host studies remain essential for detailed studies of molecular and cellular processes of host-virus interactions, culture-independent studies have greatly expanded our knowledge of microbial and viral diversity. Sequencing of the viral communities within a natural environment (termed "metaviromics") has significantly impacted environmental virology to the point where many of the most abundant viruses in the human gut (10), the ocean (11), and other natural environments (12, 13) have been described only through sequence data without isolation of the virion and the viral host. Recently, the International Committee on Taxonomy of Viruses (ICTV) decided to accept viral genome information alone as the basis for identification of a new virus (14). A set of standards for the publication of uncultivated viruses has recently been developed (15).

Following the minimum information about an uncultivated virus genome (MIUViG) guidelines (15), we present here the first viral genome that is predicted to infect a member of the Nanoarchaeota and only the second virus described to infect a member of the DPANN superphylum (16). This obligate three-way interaction among a virus, its Nanoarchaeota viral host, and the *Sulfolobales* cellular symbiont of the Nanoarchaeota represents a new system for the investigation of complex symbiotic relationships.

RESULTS

Viral genome. Viral contigs were independently assembled from both CsCl-purified and total viral metagenomes from the 1 February 2016 Crater Hills 09 (CH09) sample with metaSPAdes. After assembly, the contigs from each viral metagenome were examined and two contigs were found to be of the same length, with staggered homologous overlapping regions. Upon manual inspection of these two contigs, it was determined that their 5′ and 3′ ends were homologous to each other, indicating that the viral genome is circular. Circularization of the genome created a 35,629-bp circular double-stranded DNA (dsDNA) genome. The circular nature of the genome was confirmed by PCR amplification across the junction, by Sanger DNA sequencing of the product, and by recruitment of reads from the two viromes across the circularization junction. The assembled genome had a mean coverage depth of 435×, providing strong support for the genome sequence described. The genome was confirmed to be viral or likely viral by VirSorter analysis (17). The assembled genome had a GC content level of 27.5%. An origin of replication was not identified, likely due the high AT content of the genome. We propose that the 35.6-kb genome represents a new archaeal virus, tentatively named Nanoarchaeota Virus 1 (NAV1).

The final assembled NAV1 genome codes for 52 predicted open reading frames (ORFs) with predicted masses ranging from 5 to 146 kDa, resulting in 97% genome coding density (Fig. 1; see also Table S4 in the supplemental material). Among these

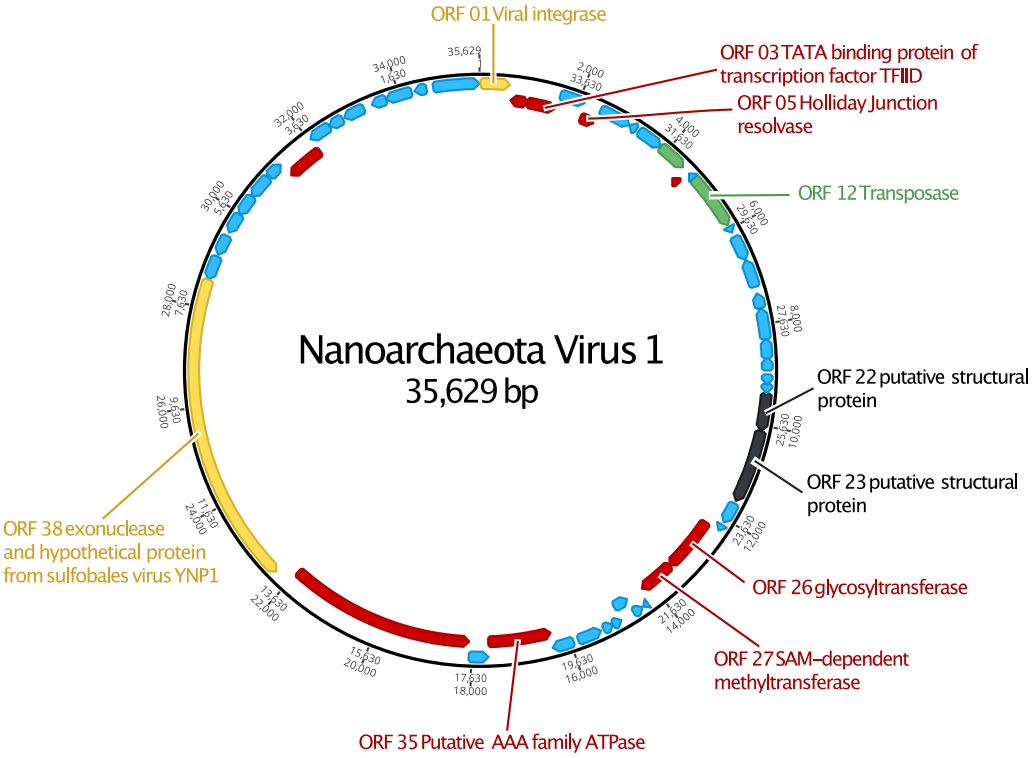

FIG 1 Genome map of NAV1. Genes are color coded based on their best BLASTx match in the NCBI nr database. Red, similar to host Nanoarchaeota genes (9 NAV1 ORFs); yellow, similar to genes of other archaeal viruses (2 NAV1 ORFs); green, similar to other archaeal species (2 NAV1 ORFs); blue, no significant similarity (37 NAV1 ORFs); black, putative structural proteins based on HMM analysis (2 ORFs). Genes with a predicted function are labeled.

ORFs, a majority (38 ORFs) had no significant similarity by BLASTn or BLASTx analysis to other sequences in the NCBI nr database (release version 228). All ORFs are described in Table S4, including the best BLASTx or hidden Markov modeling (HMM) match when available and predicted function. A total of 14 ORFs were found to show similarity to a previously characterized archaeal virus (ORF01, ORF23, and ORF38), a Nanoarchaeota gene (9 ORFs), or a gene found in another archaeal cellular species (2 ORFs). While most of the ORFs displayed no significant similarity to genes in the NCBI nr database, there were several for which we were able to predict a function (discussed below).

Further analysis of the viral genes and genome with vContact 2.0 revealed no genome-level relationships between this genome and other known viral genomes. This lack of homology between the NAV1 genome and other viral genomes strongly suggests that this virus is the founding member of a new archaeal viral family.

Fractions of virus density gradients containing the NAV1 genome were visualized by transmission electron microscopy (TEM) for the presence of virus-like particles (VLPs). TEM analysis of the CsCl fractions that contained the NAV1 genome contained VLPs with five different morphologies, including two spheres (~20 and ~60 nm in diameter), an ~400-nm-by-20-nm rod-shaped virus, and two spindle-shaped viruses with tails (Fig. 2). Due to the difficulty of packaging a 35-kb genome within a 20-nm-diameter spherical particle, it is unlikely that the 20-nm particle represents the virion associated with the NAV1 genome. Further analysis of predicted NAV1 genes by comparison to the pVOG profiles with HMM identified two adjacent ORFs that are predicted to code for minor structural proteins. One these proteins, encoded by ORF23, had a significant BLASTx match to a hypothetical protein in archaeal rod-shaped viruses Acidianus filamentous virus and *Sulfolobus islandicus* rod-shaped virus 1 (SIRV-1) to SIRV-5 and SIRV-8 to SIRV-11 (1E−8). However, what role these proteins play in these viruses is unknown. In addition, the rod-shaped VLPs observed in the electron micrographs have dimensions quite distinct from those of the two known archaeal rod-shaped viruses SIRV

FIG 2 CH09 hot spring (pH 2; 75°C) from which NAV1 was initially sequenced. The inset shows an electron microscope micrograph of possible NAV1 morphologies from a CsCl density gradient. White arrows indicate morphologies described in the text. Bar, 200 nm.

(830 by 23 nm) and APBV1 (143 by 16 nm). Likewise, the large spindle-shaped VLPs seen in the micrographs are similar to virions of Acidianus tailed spindle virus, a common virus found in the same YNP hot spring (18). However, the NAV1 genome shares no similarity with the ATSV genome. Given these observations, we suspect the NAV1 virion represents either the 60-nm spherical VLP or the 20-by-400-nm rod-shaped VLP.

Viral gene prediction. Of the 52 genes predicted in the NAV1 genome, only 10 have a predicted function, with another 5 (ORF02, ORF09, ORF10, ORF37, and ORF45) having uncharacterized homologues in archaeal species. One gene (ORF01) shows similarity to a tyrosine recombinase type viral integrase (6E−9), with 48% similarity to the integrase in *Sulfolobus tengchongensis* spindle-shaped virus 1 and 2. HHPred alignment (19) of the integrase with the XerA recombinase from *Thermoplasma acidophilum* (20) identified all four conserved residues involved in DNA-binding/protonation (Arg65, Ile90, His138, and Arg141) as well as the two catalytic tyrosine residues (Tyr172 and Tyr173). The detection of an integrase suggests that NAV1 is capable of integrating into its host cell genome. In support of this, a 3-kb segment of this viral genome flanked by cellular sequences on one end was detected in the previously sequenced Nanoarchaeota genome (8).

The presence of two large ORFs in the NAV1 genome is interesting. ORF38 codes for a 232-kDa protein (2,082 amino acids), and ORF37 codes for a 146-kDa protein (1,311 amino acids). A portion of ORF38 has two significant BLASTx matches to two ORFs found in the archaeal virus *Sulfolobales* virus YNP1. The first of these matches is to a putative exonuclease, while the second is to a hypothetical protein. However, these two matches cover only ~35% of the predicted protein of ORF38. A small region of ORF37 matches a hypothetical gene from a Nanoarchaeota cellular species along 30% of its length.

Another NAV1 gene with a putative function is a transposase encoded by ORF12. This transposase is most similar to a transposase that has previously been described in a *Sulfolobus islandicus* strain, which is part of the same family as the Nanoarchaeota host. Whether this transposase was acquired from the *Sulfolobales* symbiont or the Nanoarchaeota host remains to be determined. Transposases have also been described in other archaeal viruses and appear to be common features of archaeal viruses.

NAV1 ORF26 was predicted via blastx (E value = 0) to be a glycosyltransferase. Glycotransferases have previously been described in archaeal viruses (21), and many archaeal viruses have glycosylated capsid proteins (22, 23). Protein glycosylation is a common strategy of archaeal viruses (24) and is likely involved in enhancing virion stability in thermal environments and interactions with the cell surface (25), which is heavily glycosylated (26), or both.

NAV1 ORF27 shows high levels of homology to an *S*-adenosylmethionine (SAM)-dependent methyltransferase found in YNP Nanoarchaeota species (2E−110). Methyltransferases have previously been described in archaeal viruses (27), including SAM-dependent methyltransferases in thermophilic archaeal viruses infecting members of the *Sulfolobales* (28), suggesting that these proteins are common and widespread in high-temperature archaeal viruses.

NAV1 ORF03 is predicted to be a TATA binding protein (TBP) (1E−41 [BLASTtx match to TBP in "*Candidatus* Nanobsidianus stetteri"]) (9), the first protein to bind during transcription initiation (29) and part of the transcription factor IID (TFIID) complex. Further supporting this functional assignment, predicted TATA boxes were found upstream of numerous ORFs in the NAV1 genome.

NAV1 ORF22 and ORF23 were annotated as putative structural proteins, and their genes are adjacent to each other in the NAV1 genome (Fig. 1). The clustering of structural proteins is a common feature of viruses that facilitates synthesis of the correct number of copies of structural proteins for virion assembly. While we were not able to confirm the role of these proteins, the data provide an attractive starting point for future biochemical studies.

Virus host prediction. Multiple analyses indicate that a Nanoarchaeota species is a host for NAV1. One of the most robust methods for the prediction of viral host species is by comparison of viral genes with cellular genes (30). A BLASTn analysis comparing predicted genes in the NAV1 genome with the cellular genes in the NR database was performed. The best BLASTn match was between NAV1 ORF26 (E value 0) and a glycosyltransferase found in a YNP Nanoarchaeota species, "*Ca*. Nanobsidianus stetteri," sequenced from the Nymph Lake 01 (NL01) hot spring, and "*Ca*. Nanopusillus acidilobi," isolated from Cistern Spring in YNP (5). In addition, there were six more NAV1 genes with best BLASTn or BLASTtx matches to genes found in "*Ca*. Nanobsidianus stetteri" and two more with the best match in "*Ca*. Nanopusillus acidilobi" (Table S4). Hexanucleotide analysis was used as a second approach to link the NAV1 genome to its cellular host species. It has been demonstrated that viral genomes tend to have nucleotide frequencies similar to those of their cellular hosts (31, 32). K-mer distances of <0.3 have been demonstrated to be reliable (~40% accurate) indicators of virus-host pairs at the genus level, with increasing confidence at higher taxonomic levels (~80% accuracy at the class level [31]). Comparison of the NAV1 genome with multiple cellular metagenomic bins created from multiple YNP hot spring cellular communities indicated the smallest distance between the viral genome and a metagenomic bin, identified as a Nanoarchaeota species in both the NL01 and CH09 hot springs. The distance between these metagenomic bins and the viral genome was <0.3 for both hot springs (0.210 for CH09 and 0.239 for NL01) (Table S2), and no other bin had a distance value of <0.3 to the viral genome. Similarly, host-virus pairs tend to have similar genome GC contents. The NAV1 genome has GC content of 27.5% and the Nanoarchaeota spp. in the YNP hot springs have GC content of ~25%, while all other cellular species in the NL01 and CH09 hot springs have GC content of >34%. The final evidence that this virus infects a Nanoarchaeota species comes from single-cell genomic studies from the NL01 hot spring (33). In this study, over 300 single cells from the NL01 hot spring were sequenced, and 250 cells were identified to the species level. Six of these single cells were identified to be Nanoarchaeota, e.g., Nanoarchaeota cells without a host cell associated with them. Sequence composition analysis of these 250 single cells revealed that the 6 Nanoarchaeota cells had the smallest k-mer usage distance to the NAV1 genome than any of other 245 non-Nanoarchaeota cells in the data set (Table S2). In addition to the

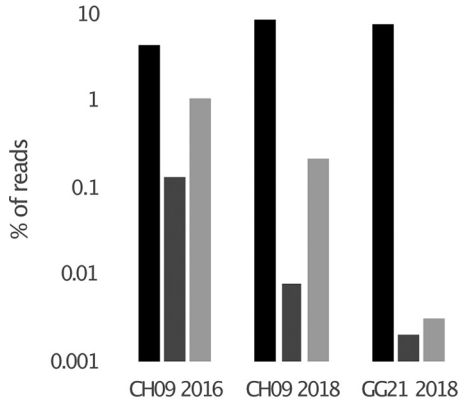

FIG 3 Virus and host relative abundances indicated by read recruitment in three hot spring samples where a cellular metagenome bin was identified as Nanoarchaeota. Relative levels of host (black) and viral (dark gray) abundance in three cellular metagenomes and viral abundance in free viral metaviromes (light gray) from the same hot spring and time point are indicated.

single sorted Nanoarchaeota cells there were 7 additional cosorted Nanoarchaeota cells, e.g., Nanoarchaeota cells with a host cell that was cosorted with the Nanoarchaeota cell (7, 33). Five of these cosorted cells had sequence reads that map to the NAV1 genome. None of the other single cells had sequence reads that mapped to the NAV1 genome. In addition, NAV1 viral sequences were detected within a Nanoarchaeota single cell. Mauve based alignment of this cellular genomic region with the NAV1 genome sequence revealed a 3 kb region with homology between the viral genome and the single-cell genome. The low (~30%) level of homology suggests that the viral sequence represents a remnant of a past infection. Taken together, these multiple lines of evidence, including data from BLAST analysis, k-mer analysis, and matches to sequences present in single-cell genomes, indicate that Nanoarchaeota spp. are a host for NAV1.

Distribution and relative abundances of NAV 1 across Yellowstone hot springs. It has previously been demonstrated that the Nanoarchaeota are widely distributed across YNP high-temperature low-pH hot springs (8). In order to examine the distribution of NAV1 across YNP hot springs, we compared the distribution of its Nanoarchaeota host with the presence of viral sequences. Sequence reads from viral and cellular metagenomes generated from 5 YNP hot springs were recruited to the NAV1 genome. While reads corresponding to the NAV1 genome were detected in all 5 hot springs, significantly more reads were detected in 2 of the hot springs where metagenomic bins corresponding to the Nanoarchaeota hosts were also identified (CH09 and Gibbon Geyser 21 [GG21]) (Fig. 3), indicating that the virus is more abundant in these hot springs. The largest relative abundance of NAV1 was seen in the 1 February 2016 CH09 sample from which the virus was originally sequenced. At that time, 63,127 reads (1.1%) from the viral fraction and 18,107 reads (0.13%) from the cellular metagenome were recruited across the entire NAV1 genome. In the 2018 CH09 sample, 10,731 reads (0.2%) were recruited from the viral metagenome across 86% of the NAV1 genome and 2,238 reads (0.08%) from the cellular metagenome. In the NL01 hot spring sampled in July 2018, where the abundance of the Nanoarchaeota was significantly reduced from that seen at previous sampling times, only 0.016% of viral reads and 0.075% of cellular reads were recruited to 25% of the NAV1 genome. These differences between the relative abundances of the virus in different hot springs and across time suggest that, while the population is stable in the community, the absolute abundance of the virus changes over time.

DISCUSSION

While it is commonly assumed that viruses are capable of infecting all types of cells, whether the smallest and most divergent cells, such as the Nanoarchaeota and *Rick-*

ettsia, are capable of supporting viral replication has remained an unresolved issue. Here, we present the complete viral genome of NAV1, an archaeal virus that infects a Nanoarchaeota spp. (Fig. 1). This is the first virus described to infect a Nanoarchaeota species. While this virus remains uncultivated, it meets the standards for identification of a new virus by genome sequence alone (15) (see Table S5 in the supplemental material). NAV1 is a common member of viral communities present in YNP acidic hot springs. It also represents one of the first examples of three-way interaction between a virus infecting a single-cell species that is itself an ectobiont of different single-cell species. This is the first identification of such a three-way interaction in the domain Archaea.

The role that NAV1 plays in the three-way symbiotic interaction is unknown. While the two-way cellular interactions between Nanoarchaeota and its host cell are poorly understood, recent studies have suggested that there is an energy transfer from the host cell to the Nanoarchaeota (7). It remains to be determined if NAV1 receives any necessary components for its replication from the cellular host of Nanoarchaeota. It is likely that this is the case, since the Nanoarchaeota themselves lack most of the genes required for the synthesis of amino acids and nucleotides, essential components for viral replication. While the physical site of Nanoarchaeota virus replication is unknown, we suspect that it is within in the Nanoarchaeota cell and not within the host cell of Nanoarchaeota on the basis of our inability to detect viral sequences with single cells of *Acidicryptum* spp., a host cell of YNP Nanoarchaeota. However, it remains to be determined how either the suspected 60-nm-diameter spherical virus or 400-nm-long rod-shaped virus particles could occupy an approximately 300-nm-diameter Nanoarchaeota host cell.

While we were unable to positively determine the morphology of NAV1, we have narrowed the likely morphologies down to two virion morphologies. We suspect that the NAV1 virion has either a 60-nm spherical or a 400-by-40-nm rod-shaped virion morphology on the basis of its genome size and genome content. The lack of NAV1 genome similarity to other known viruses suggests that this virus is related only distantly to other archaeal viruses. Previous efforts at characterizing the archaeal virosphere have demonstrated poor connectivity between archaeal viral groups with few exceptions and indicated that most of the viral groups are evolutionarily distinct from each other (34). Similarly to other archaeal viruses (35), a majority of the genes carried by NAV1 cannot be functionally annotated by sequence analysis alone and most (37/52) show no significant similarity via blastn or blastp in the nr database. While NAV1 is more similar to other archaeal viruses than to bacteriophage, its genomic differences are sufficient that, while the final classification of this virus will likely require confirmation of the viral morphology, it is likely that this virus is will be the founding member of a new archaeal viral family. It is tempting to hypothesize that this distinction from other archaeal viruses has arisen due to the coevolution of NAV1 and its Nanoarchaeota host.

The establishment of a culture-based system for NAV1 and its Nanoarchaeota host will provide critical insight into the exchange of metabolites between its cellular symbiont, Nanoarchaeota, and the virus. It will be interesting to determine the details of the NAV1 replication cycle and the impact of viral replication on this 3-way symbiosis.

MATERIALS AND METHODS

Viral sampling, detection, and enrichment. Hot spring water samples were initially collected from the Crater Hills 09 (CH09) hot spring (see Table S1 in the supplemental material). All samples were filtered through 0.2-um-pore-size in-line filters, and the virus particles in the flow through were further concentrated using the $FeCl_2$ flocculation method (36). Due to the high levels of iron already present in these hot springs, no additional iron was added. The pH was raised to pH 4.5 to precipitate viruses. After precipitation, the virus-FeCl precipitates were collected on 0.2-um-pore-size filters and resuspended in 500 mM ascorbic acid (pH 3). After resuspension, the viral concentrate was dialyzed against 5 mM citrate buffer (pH 3) three times for 30 min each time. The virus was further concentrated using a 100,000-molecular-weight-cutoff (MWCO) Microcon spin column (Millpore Sigma, Darmstadt Germany) at 10,000 $\times$ *g* to reach a final volume of 500 μl and was subsequently treated with RQ DNAse1 following the protocol of the manufacturer (Promega, Madison, WI). Viral DNA was extracted from virus particles using an AllPrep PowerViral DNA/RNA kit (Qiagen, Hilden Germany). Virus from CH09 collected on 1 February 2016 were further purified by banding on 1.35-g/ml CsCl density gradients generated by

centrifugation at 38,000 rpm for 48 h at 10°C in a SW41 rotor. After centrifugation, the gradient was fractionated and the location of archaeal viral sequence determined using quantitative PCR (qPCR) and previously described PCR primers (8) to identify the virus. The virus was located within the 1.33-g/ml gradient fraction. This fraction was treated with Rq DNase 1 before DNA extraction was performed with the AllPrep PowerViral DNA/RNA kit. Libraries were created from extracted DNA and sequenced at the University of Illinois Genomics Center on an Illumina MiSeq system using V2 chemistry.

After identification of the viral genome, additional samples were collected from the CH09 hot spring as well as from Nymph Lake 01 (NL01), Gibbon Geyser 20 (GG20), and GG21, all high-temperature (>80°C) acidic (pH <4) hot springs in YNP, and from GG08, a low-temperature acidic hot spring in YNP, to examine viral distribution in the hot springs of YNP (Table S1). All samples were processed as described above with the exception that CsCl density purification was not performed prior to viral DNA (vDNA) extraction.

Genome assembly and identification of virion. Sequence reads were assembled with metaSPAdes using the default parameters (37). After assembly, the genome was circularized by manual examination of the genome. The genome sequence was confirmed by PCR amplification of genome segments and Sanger sequencing. The genome was examined with VirSorter 1.0.3 (17) in decontamination mode using the virome database in the Discovery Environment on the CyVerse infrastructure. CsCl-purified virus was stained with uranyl acetate and imaged with a Leo 912 transmission electron microscope.

Gene prediction. Open reading frames (ORFs) were predicted with Glimmer (38) in combination with Geneious (V10.2.5) using a six-frame translation of the viral genome. Manually verified and predicted genes were annotated using a combination of BLASTn, BLASTx, and the prokaryotic Viral Orthologous Groups (pVOG) profiles (39), as well as vContact for determinations of genome-level and gene-level relationships with other viruses (40).

Host prediction. Virus-host nucleotide pairings were calculated for the Nanoarchaeota virus genome against assembled cellular contigs bins from the CH09 and NL01 hot springs (J. H. Munson-McGee and M. Young, unpublished data), using the d2* method in VirHostMatcher (31) (Table S2). Additional nucleotide distance usages were calculated using 15 high-coverage single-cell genomes from NL01 (8) and nearly 300 low-coverage single-cell genomes from the NL01 hot spring (33) representing the eight dominant cellular species present in the NL01 hot spring (Table S2).

Three hundred previously sequenced low-coverage single-cell genomes from the same NL01 hot spring (33) were also used to identify cells containing the Nanoarchaeota virus genome sequence. Sequence reads from these single cells were recruited to the Nanoarchaeota virus genome using Bowtie2 (41). Recruited sequence reads were required to be 98% identical over a minimum of 100 bp in order for a read to be identified.

Virus distribution and relative abundances. Sequence reads from viral and cellular metagenomes generated from CH09 and four additional hot springs (Table S1) collected approximately 38 months after initial sampling were recruited to the assembled Nanoarchaeota virus genome with Bowtie2 using the high-sensitivity end-to-end default settings. Reads from the cellular metagenomes were also recruited to the Nanoarchaeota single-cell AB-777-O03 genome (8) using the same settings (Table S3). The percentage of reads recruited to the viral and cellular genomes from each hot spring was calculated and used as a measure of relative abundances of the Nanoarchaeota host and viral genome in each hot spring.

Data availability. The metagenomic reads used for assembly are available at the Sequence Read Archive (SRA) under accession number PRJNA579086. The complete sequence of NAV1 is available under GenBank accession number MN056826.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank Jennifer Wirth, Ross Hartman, Cassia Wagner, and Sue Brumfield for their advice and support.

This work was supported by NSF grant DEB-4W4596 to M.J.Y.

This work was conducted in Yellowstone National Park under the terms of permits YELL-2016-SCI-5090, YELL-2017-SCI-5090, and YELL-2018-SCI-5090.

We declare that we have no conflict of interest.

REFERENCES

1. Althauser M, Samsonoff WA, Anderson C, Conti SF. 1972. Isolation and preliminary characterization of bacteriophages for *Bdellovibrio bacteriovorus*. J Virol 10:516–523.
2. Varon M, Levisohn R. 1972. Three-membered parasitic system: a bacteriophage, Bdellovibrio bacteriovorus, and Escherichia coli. J Virol 9:519–525.
3. Markelova NY. 2010. Predacious bacteria, Bdellovibrio with potential for biocontrol. Int J Hyg Environ Health 213:428–431. https://doi.org/10.1016/j.ijheh.2010.08.004.
4. Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417:63–67. https://doi.org/10.1038/417063a.
5. Wurch L, Giannone RJ, Belisle BS, Swift C, Utturkar S, Hettich RL, Reysenbach A-L, Podar M. 2016. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. Nat Commun 7:12115. https://doi.org/10.1038/ncomms12115.

6. St JE, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, Lagutin K, Mitchell K, Reysenbach A-L. 16 August 2018, posting date. A new symbiotic Nanoarchaeote (Candidatus Nanoclepta minutus) and its host (Zestosphaera tikiterensis gen. nov., sp. nov.) from a New Zealand hot spring. Syst Appl Microbiol https://doi.org/10.1016/j.syapm.2018.08.005.
7. Jarett JK, Nayfach S, Podar M, Inskeep W, Ivanova NN, Munson-McGee J, Schulz F, Young M, Jay ZJ, Beam JP, Kyrpides NC, Malmstrom RR, Stepanauskas R, Woyke T. 2018. Single-cell genomics of co-sorted Nanoarchaeota suggests novel putative host associations and diversification of proteins involved in symbiosis. Microbiome 6:161. https://doi.org/10.1186/s40168-018-0539-8.
8. Munson-McGee JH, Field EK, Bateson M, Rooney C, Stepanauskas R, Young MJ. 2015. Nanoarchaeota, their Sulfolobales host, and Nanoarchaeota virus distribution across Yellowstone National Park hot springs. Appl Environ Microbiol 81:7860–7868. https://doi.org/10.1128/AEM.01539-15.
9. Podar M, Makarova KS, Graham DE, Wolf YI, Koonin EV, Reysenbach A-L. 2013. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biol Direct 8:9. https://doi.org/10.1186/1745-6150-8-9.
10. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 5:4498. https://doi.org/10.1038/ncomms5498.
11. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Patterns and ecological drivers of ocean viral communities. Science 348:1261498–1261410. https://doi.org/10.1126/science.1261498.
12. Bolduc B, Wirth JF, Mazurie A, Young MJ. 2015. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. ISME J 9:2162–2177. https://doi.org/10.1038/ismej.2015.28.
13. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li C, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol 3:870–880. https://doi.org/10.1038/s41564-018-0190-y.
14. Simmonds P, Adams MJ, Benk M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, Van Der Vlugt RA, Varsani A, Zerbini FM. 2017. Consensus statement: virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15:161–168. https://doi.org/10.1038/nrmicro.2016.177.
15. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, et al. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol 37:29–37. https://doi.org/10.1038/nbt.4306.
16. Martínez-García M, Santos F, Moreno-Paz M, Parro V, Antón J. 2014. Unveiling viral–host interactions within the 'microbial dark matter'. Nat Commun 5:4542. https://doi.org/10.1038/ncomms5542.
17. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.
18. Hochstein R, Amenabar MJ, Munson-McGee JH, Boyd ES, Young MJ. 2016. Acidianus tailed spindle virus: a new archaeal large tailed spindle virus discovered by culture-independent methods. J Virol 90:3458–3468. https://doi.org/10.1128/JVI.03098-15.
19. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a New HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.
20. Han A, Hwang KY, Park SY, Kim J, Nam KH, Jo CH. 2016. Crystal structure of Thermoplasma acidophilum XerA recombinase shows large C-shape clamp conformation and cis-cleavage mode for nucleophilic tyrosine. FEBS Lett 590:848–856. https://doi.org/10.1002/1873-3468.12109.
21. Larson ET, Reiter D, Young M, Lawrence CM. 2006. Structure of A197 from Sulfolobus turreted icosahedral virus: a crenarchaeal viral glycosyltransferase exhibiting the GT-A fold. J Virol 80:7636–7644. https://doi.org/10.1128/JVI.00567-06.
22. Quemin ERJ, Pietilä MK, Oksanen HM, Forterre P, Rijpstra WIC, Schouten S, Bamford DH, Prangishvili D, Krupovic M. 2015. Sulfolobus spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein and host-derived lipids. J Virol 80:e02270-15. https://doi.org/10.1128/JVI.02270-15.
23. Mochizuki T, Yoshida T, Tanaka R, Forterre P, Sako Y, Prangishvili D. 2010. Diversity of viruses of the hyperthermophilic archaeal genus Aeropyrum, and isolation of the Aeropyrum pernix bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. Virology 402:347–354. https://doi.org/10.1016/j.virol.2010.03.046.
24. Maaty WS, Ortmann AC, Dlakic M, Schulstad K, Hilmer JK, Liepold L, Weidenheft B, Khayat R, Douglas T, Young MJ, Bothner B. 2006. Characterization of the archaeal thermophile Sulfolobus turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. J Virol 80:7625–7635. https://doi.org/10.1128/JVI.00522-06.
25. Markine-Goriaynoff N, Gillet L, Van Etten JL, Korres H, Verma N, Vanderplasschen A. 2004. Glycosyltransferases encoded by viruses. J Gen Virol 85:2741–2754. https://doi.org/10.1099/vir.0.80320-0.
26. Quemin ERJ, Quax T. 2015. Archaeal viruses at the cell envelope: entry and egress. Front Microbiol 6:552. https://doi.org/10.3389/fmicb.2015.00552.
27. Baranyi U, Klein R, Lubitz W, Krüger DH, Witte A. 2000. The archaeal halophilic virus-encoded Dam-like methyltransferase M.φCh1-I methylates adenine residues and complements dam mutants in the low salt environment of Escherichia coli. Mol Microbiol 35:1168–1179. https://doi.org/10.1046/j.1365-2958.2000.01786.x.
28. Liu Y, Ishino S, Ishino Y, Pehau-Arnaudet G, Krupovic M, Prangishvili D. 2017. A novel type of polyhedral viruses infecting hyperthermophilic Archaea. J Virol 91:e00589-17. https://doi.org/10.1128/JVI.00589-17.
29. Kawakami E, Adachi N, Senda T, Horikoshi M. 2017. Leading role of TBP in the establishment of complexity in eukaryotic transcription initiation systems. Cell Rep 21:3941–3956. https://doi.org/10.1016/j.celrep.2017.12.034.
30. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev 40:258–272. https://doi.org/10.1093/femsre/fuv048.
31. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. 2017. Alignment-free d*2 oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res 45:39–53. https://doi.org/10.1093/nar/gkw1002.
32. Roux S, Hallam SJ, Woyke T, Sullivan MB. 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife 4:1–20. https://doi.org/10.7554/eLife.08490.
33. Munson-McGee JH, Peng S, Dewerff S, Stepanauskas R, Whitaker RJ, Weitz JS, Young MJ. 2018. A virus or more in (nearly) every cell: ubiquitous networks of virus–host interactions in extreme environments. ISME J 12:1706–1714. https://doi.org/10.1038/s41396-018-0071-7.
34. Iranzo J, Koonin EV, Prangishvili D, Krupovic M. 2016. Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. J Virol 90:11043–11055. https://doi.org/10.1128/JVI.01622-16.
35. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV. 2018. Viruses of archaea: structural, functional, environmental and evolutionary genomics. Virus Res 244:181–193. https://doi.org/10.1016/j.virusres.2017.11.025.
36. John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environ Microbiol Rep 3:195–202. https://doi.org/10.1111/j.1758-2229.2010.00208.x.
37. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 1:30–47.

38. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641. https://doi.org/10.1093/nar/27.23.4636.
39. Grazziotin AL, Koonin EV, Kristensen DM. 2017. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res 45:D491–D498. https://doi.org/10.1093/nar/gkw975.
40. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol 37:632–639. https://doi.org/10.1038/s41587-019-0100-8.
41. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.