10-30-2019

An Uncultivated Virus Infecting a Symbiotic Nanoarchaeota in the Hot Springs of Yellowstone National Park

Jacob H. Munson-McGee
Montana State University - Bozeman

Colleen Rooney
Portland State University

Mark J. Young
Montana State University - Bozeman

Citation Details

Munson-McGee, J. H., Rooney, C., & Young, M. J. (2019). An uncultivated virus infecting a symbiotic Nanoarchaeota in the hot springs of Yellowstone National Park. Journal of virology. DOI 10.1128/JVI.01213-19

An uncultivated virus infecting a symbiotic Nanoarchaeota in the hot springs of Yellowstone National Park

Jacob H. Munson-McGee,[1]* Colleen Rooney,[2]* Mark J. Young[3]#

Running Title: An uncultivated Nanoarchaeota virus

Author affiliations
[1]Department of Microbiology and Immunology, Montana State University, Bozeman, Montana, USA
[2]Department of Chemistry and Biochemistry, Montana State University, Bozeman, Montana, USA
[3]Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, Montana, USA

*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.

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

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 have been described that infect the Nanoarchaeota. Here we identify a virus infecting Nanoarchaeota using a combination of viral metagenomic and bioinformatic approaches. This virus, tentatively named Nanoarchaeota Virus 1 (NAV1), consists of a 35.6kb circular DNA genome encoding 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 for the first virus 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. Based on 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

Introduction

Viruses are ubiquitous features of cellular life. Their discovery and characterization have often led to new insights into the biology of cells in which they infect. Most of the viruses that infect parasites are bacteriophage that infect enodosymbiotic or disease-causing bacteria. One of 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, increasing the phage titer (2). However, the interaction between the *Bdellovibrio* and their prey as well as the *Bdellovibrio* and their phage are more akin to classic predator prey relationships than a long-term parasitic relationship.

The Nanoarchaeota are 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 interests due to their highly reduced genomes, rapidly evolving gene sequences, symbiotic lifestyle, and phylogenetic placement within the Archaea. Originally described as a deep branching phylum of the Archaea, more recent analyses have classified the Nanoarchaeota as part of the DPANN archaeal superphylum. Despite ribosomal RNA surveys suggesting that Nanoarchaeota are widespread in marine and terrestrial environments, there are currently only three cultured representatives of the Nanoarchaeota. The original *Nanoarchaeam equitans* isolated from a shallow marine hydrothermal vent, (4) as

well as the terrestrial *Nanopusillus acidilobi* isolated from a circumneutral hot spring in Yellowstone National Park (YNP) (5), and the recently described Candidatus *Nanoclepta minutus* and its host *Zestosphaera tikiterensis* from a hot spring in New Zealand (6). In addition, there multiple examples of terrestrial Nanoarchaeota genomes that have been sequenced but remain uncultured (7–9). A recent analysis of Nanoarchaeota and their co-sorted 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 only been described 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 described virus to infect a member of the DPANN superphylum (16). This obligate three-way interaction between virus, its Nanoarchaeota viral host, and the Sulfolobales cellular symbiont of the Nanoarchaeota describes a new system for the investigation of complex symbiotic relationships.

Materials and Methods

Viral Sampling, Detection, and Enrichment

Hot spring water samples were initially collected from the Crater Hills 09 (CH09) hot spring (Supplemental Table 1). All samples were filtered through 0.2um in-line filters and the virus particles in the flow through were further concentrated using the $FeCl_2$ flocculation method (17). 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 0.2um filters and resuspended in 500mM ascorbic acid pH 3. After resuspension, the viral concentrate was dialyzed against 5mM citrate buffer pH 3 for 3 x 30 min. The virus was further concentrated using a 100,000 MWCO Microcon spin column (Millpore Sigma, Darmstadt Germany) at 10,000 x g to a final volume of 500ul which was subsequently treated with RQ DNAse1 following the manufacturer's protocol (Promega, Madison, WI). Viral DNA was extracted from virus particles using the AllPrepPowerViral DNA/RNA kit (Qiagen, Hilden Germany). Virus from CH09 collected on 2016-02-01 were further purified by banding on 1.35g/ ml CsCl density gradients generated by centrifugation at 38,000rpm for 48 hours, 10˚C, in a SW41 rotor. After centrifugation, the gradient was

fractionated and the location of archaeal viral sequence determined qPCR using the PCR primers previously described to identify the virus (8). The virus was located within the 1.33g/ml gradient fraction. This fraction was treated with Rq DNAse 1 before DNA extraction with the AllPrepPowerViral DNA/RNA kit. Libraries were created from extracted DNA and sequenced at the University of Illinois Genomics Center on an Illumina MiSeq with V2 chemistry.

After identification of the viral genome, additional samples were collected from the CH09 hot spring as well as the Nymph Lake 01 (NL01), Gibbon Geyser 20, and Gibbon Geyser 21 (GG21), all high temperature (>80C) acidic (pH<4) hot springs and Gibbon Geyser 08 (GG08) a low temperature acidic hot spring in YNP to examine viral distribution in the hot springs of YNP (Supplemental Table 1). All samples were processed as described above with the exception that CsCl density purification was not performed prior to vDNA extraction.

Genome Assembly and Identification of Virion

Sequence reads were assembled with metaSPAdes using the default parameters (18). 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 (19) 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

Journal of Virology

Open reading frames (ORfs) were predicted with Glimmer (20) 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 (21), as well as vContact for genome level and gene level relationships with other viruses (22).

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 (Munson-McGee et al in prep), using the d2* method in VirHostMatcher (23) (Supplemental Table 2). Additional nucleotide distance usages were calculated using fifteen high coverage single cell genomes from NL01 (8), and nearly 300 low coverage single cell genomes from the NL01 hot spring (24) representing the eight dominant cellular species present in the NL01 hot spring (Supplemental Table 2).

Three hundred previously sequenced low coverage single cell genomes from the same NL01 hot spring (24) were also used to identify cells containing the Nanoarchaeota virus genome sequence. Sequence reads from these single cells were recruited onto the Nanoarchaeota virus genome using Bowtie2 (25). Recruited sequence reads were required to be 98% identical over a minimum of 100bp in order for a read to be identified.

Virus Distribution and Relative Abundance

Sequence reads from viral and cellular metagenomes generated from CH09 and four additional hot springs (Supplemental Table 1) collected approximately 38 months after initial

168 sampling were recruited to the assembled Nanoarchaeota virus genome with Bowtie2 using the
169 high sensitivity end-to-end default settings. Reads from the cellular metagenomes were also
170 recruited to the Nanoarchaeota single cell genome AB-777-O03 (8) using the same settings
171 (Supplemental Table 3). The percent of reads recruited to the viral and cellular genomes from
172 each hot spring was calculated and used as a measure of relative abundance of the
173 Nanoarchaeota host and viral genome in each hot spring.

174 **Data Availability**

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

178

179 ## Results

180

181 **Viral Genome**

182

183 Viral contigs were independently assembled from both the CsCl purified and total viral
184 metagenomes from the February 1st 2016 CH09 sample with metaSPAdes. After assembly the
185 contigs from each viral metagenome were examined and two contigs were found to be of the
186 same length with a staggered homologous overlapping region to each other. Upon manual
187 inspection of these two contigs, it was determined that their 5' and 3' ends were homologous
188 to each other Indicating that the viral genome is circular. Circularization of the genome created
189 a 35,629 bp circular ds DNA genome. The circular nature of the genome was confirmed by PCR
190 amplification across the junction, Sanger DNA sequencing of the product, and by read
191 recruitment from the two viromes across the circularization junction. The assembled genome

had a mean coverage depth of 435x providing strong support for the genome sequence described. The genome was confirmed to be viral or likely viral by VirSorter analysis (19). The assembled genome had a GC content 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 ORFs with predicted masses ranging from 5-146 kDa, resulting in a 97% genome coding density (Figure 1, Supplemental Table 4). Of these ORFs, a majority (38 ORfs) have no significant similarity by BLASTn or BLASTx analysis to other sequences in the NCBI nr database (release version 228). All ORFs are described in Supplemental Table 4 including the best BLASTx or Hidden Markov modeling (HMM) match when available and predicted function. 14 ORFs either have similarity to a previously characterized archaeal virus (ORFs 01, 23, and 38), a Nanoarchaeota gene (9 OFRs) 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 are several that 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 suggest that this virus is the founding member of a new archaeal viral family.

Virus density gradients fractions containing the NAV1 genome were visualized by 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, two spheres, ~20 and

~60nm in diameter, a ~400nm x 20nm rod-shaped virus and two spindle shaped viruses with tails (Figure 2). Due to the difficulty in packaging a 35kb genome within a 20nm diameter spherical particle it is unlikely that the 20nm particle is 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 ORF 23, has a significant BLASTx match to a hypothetical protein in archaeal rod-shaped viruses Acidianus filamentous virus and SIRVs 1-5 and 8-11 (1E-8). However, it is unknown what role these proteins play in these viruses. In addition, the rod-shaped VLPs observed in the electron micrographs have dimensions quite distinct from the two known archaeal rod shaped viruses SIRV (830 x 23 nm) and APBV1 (143 x 16nm). Likewise, the large spindled-shaped VLP seen in the micrographs are similar to ATSV virion, a common virus found in the same YNP hot spring (26). However, the NAV1 genome shares no similarity to the ATSV genome. Given these observations, we suspect the NAV1 virion is either the 60 nm spherical VLP or the 20 x 400nm rod-shaped VLP.

Viral Gene Prediction

Of the 52 genes predicted in the NAV1 genome, only 10 have a predicted function, with another 5 (ORFs 02, 09, 10, 37, and 45) having uncharacterized homologues in archaeal species. One gene (ORF 01) shows similarity to a tyrosine recombinase type viral integrase (6E-9), with a 48% similarity to the integrase in *Sulfolobus tengchongensis* spindle-shaped virus 1 and 2. HHPred alignment (27) of the integrase with the XerA recombinase from *Thermoplasma acidophilium* (28) identified all four conserved residues involved in DNA-binding/protonation (Arg65, Ile90, His138, and Arg141) as well as the two catalytic tyrosine residues (Tyr172,

Tyr173). The detection of an integrase suggests that NAV1 is capable of integrating into its host cell genome. In support of this, a 3kb segment of this viral genome flanked by cellular sequences on one end was detected in previous sequence Nanoarchaeota genome (8).

It is interesting to note the presence of two large ORFs in the NAV1 genome. ORF 38 codes for a 232 kDa protein (2,082 amino acids) and ORF 37 codes for 146 kDa protein (1,311 amino acids). A portion of ORF 38 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 only cover ~35% of the predicted protein of ORF 38. A small region of ORF 37 matches to a hypothetical gene from a Nanoarchaeota cellular species along 30% of the its gene length.

Another NAV1 gene with a putative function is a transposase encoded by ORF 12. This transposase is most similar to a transposase that has previously been described in a *Sulfolobus islandicus* species, which is part of the same family as the Nanoarchaeota host. Whether this transposase was acquired from the Sulfolobales symbiont or the Nanoarchaeota host still remains to be determined. Transposases have also been described in other archaeal viruses and appear to be common features of archaeal viruses.

NAV1 ORF 26 was predicted via blastx (evalue=0) to be a glycosyltransferase. Glycotransferases have previously been described in archaeal viruses (29) and many archaeal viruses have glycosylated capsid proteins (30, 31). Protein glycosylation is a common strategy of archaeal viruses (32) and is likely involved in enhancing the virion stability in thermal environments and or interactions with the cell surface (33), which is heavily glycosylated (34).

257 NAV1 ORF 27 shows high levels of homology to a SAM-dependent methyltransferase
258 found in YNP Nanoarchaeota species (2e-110). Methyltransferases have previously been
259 described in archaeal viruses (35) including SAM dependent methyltransferases in thermophilic
260 archaeal viruses infecting members of the Sulfolobales (36), suggesting that these proteins are
261 common and widespread in high-temperature archaeal viruses.
262 NAV1 ORF 03 is predicted to be the TATA binding protein (TBP)(1E-41 BLASTtx match to
263 TBP in *N. stetteri*) (9), the first protein to bind during transcription initiation (37), and part of
264 the TFIID complex. Further supporting this functional assignment predicted TATA boxes were
265 found upstream of numerous ORFs in the NAV1 genome.
266 NAV1 ORF 22 and ORF 23 were annotated as putative structural proteins and their
267 genes are adjacent to each other in the NAV1 genome (Figure 1). The clustering of structural
268 proteins is a common feature of viruses to facilitate synthesis of the correct number of copies
269 of structural proteins for virion assembly. While we were not able to confirm the role of these
270 proteins, it does provide an attractive starting point for future biochemical studies.
271
272 **Virus Host Prediction**
273
274 Multiple analyses indicate that a Nanoarchaeota species is a host for NAV1. One of the
275 most robust methods for the prediction of viral host species is by comparison of viral genes
276 with cellular genes (38). A BLASTn analysis of predicted genes in the NAV1 genome with the
277 cellular genes in the NR database was performed. The best BLASTn match was between NAV1
278 ORF 26 (E value 0) with a glycosyltransferase found in YNP Nanoarchaeota species, candidatus
279 *N. stetteri*, sequenced from the NL01 hot spring and the *Nanopusillus acidilobi* isolated from
280 Cistern Spring in YNP (5). In addition, there were six more NAV1 genes with a best BLASTn or

BLASTtx matches to genes found in a Candidatus *N. stetteri* and two more with the best match in *N. acidiliobi* (Supplemental Table 4). Hexanucleotide analysis was used as a second approach to link the NAV1 virus genome to its cellular host species. It has been demonstrated that viral genomes tend to have similar nucleotide frequencies as their cellular hosts (23, 39). K-mer distances <0.3 have been demonstrated to be reliable indicators of virus host pairs at the genus level (~40% accurate) with increasing confidence at higher taxonomic levels (~80% accuracy at the class level (23)). 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 in both hot springs, CH09 (0.210) and NL01 (0.239) (Supplemental Table 2) and no other bin had a distance <0.3 to the viral genome. Similarly, hosts-virus pairs tend to have similar genome GC content. The NAV1 genome has a GC content of 27.5% and the Nanoarchaeota sp in the YNP hot springs have a GC content of ~25% while all other cellular species in the NL01 and CH09 hot springs have a 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 (24). 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 dataset (Supplemental Table 2). In addition to the

303 single sorted Nanoarchaeota cells there were 7 additional co-sorted Nanoarchaeota cells, eg
304 Nanoarchaeota cells with a host cell that was co-sorted with the Nanoarchaeota cell (7, 24).
305 Five of these co-sorted cells had sequence reads that map to the NAV1 genome. None of the
306 other single cells had sequence reads that mapped to the NAV1 genome. In addition, NAV1 viral
307 sequences were detected within a Nanoarchaeota single cell. Mauve based alignment of this
308 cellular genomic region with the NAV1 genome sequence revealed a 3kb region with homology
309 between the viral genome and the single cell genome. The low level of homology (~30%)
310 suggests that the viral sequence is the remnant of a past infection. Taken together these
311 multiple lines of evidence, BLAST analysis, k-mer analysis, and matches to sequences present in
312 single cell genomes indicate that Nanoarchaeota spp. are a host for NAV1.
313
314 **Distribution and Relative Abundance of NAV 1 Across Yellowstone Hot Springs**
315
316 It has previously been demonstrated that the Nanoarchaeota are widely distributed
317 across YNP high temperature low pH hot springs (8). In order to examine the distribution of
318 NAV1 across YNP hot springs, we compared the distribution of its Nanoarchaeota host with the
319 presence of viral sequences. Sequence reads from viral and cellular metagenomes generated
320 from 5 YNP hot springs viral were recruited to the NAV1 genome. While reads to the NAV1
321 genome were detected in all 5 hot springs, significantly more reads were detected in 2 of the
322 hot springs where metagenomic bins corresponding to the Nanoarchaeota hosts were also
323 identified (CH09 and GG21) (Figure 3), indicating that the virus is more abundant in these hot
324 springs. The largest relative abundance of NAV1 was in the February 1st 2016 CH09 sample from
325 which the virus was originally sequenced. At this time 63,127 reads (1.1%) from the viral
326 fraction and 18,107 reads (0.13%) from the cellular metagenome recruited across the entire

NAV1 genome. In the 2018 CH09 sample 10,731 (0.2%) reads were recruited from the viral metagenome across 86% of the NAV1 genome and 2238 (0.08%) reads from the cellular metagenome. In the NL01 hot spring sampled in July 2018, were the abundance of the Nanoarchaeota was significantly reduced from 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 abundance of the virus in different hot springs and across time suggests that while 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, it has remained an unanswered question if the smallest and most divergent cells, such as the Nanoarchaeota and *Rickettsia* are capable of supporting viral replication. Here we present here the complete viral genome of NAV1, an archaeal virus that infects a Nanoarchaeota spp. (Fig. 1). This is the first described virus to infect a Nanoarchaeota. While this virus remains uncultivated, it meets the standards to identify a new virus by genome sequence alone (15) (Supplemental Table 5). NAV1 is a common member of viral communities present in YNP acidic hot springs. It is also 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.

It is unknown what role the NAV1 virus plays in the three-way symbiotic interaction. 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 Nanoarchaeota's cellular host. It likely that this is the case since the Nanoarchaeota themselves lacks most of the genes required for the synthesis of amino acids and nucleotides, essential components for viral replication. While the physical site of the Nanoarchaeota virus replication is unknown, we suspect that it is within in the Nanoarchaeota cell and not within the host cell of Nanoarchaeota based on our inability to detect viral sequences with single cells of the *Acidicryptum* spp, a host cell of YNP Nanoarchaeota. However, it remains to be determined how either of the suspected 60nm spherical or 400nm rod-shaped virus particles could occupy an approximately 300nm diameter Nanoarchaeota host cell.

While we were unable to positively identify the morphology of NAV1, we have narrowed the likely morphologies down to two virion morphologies. We suspect that the NAV1 virion is either a 60nm spherical or a 400 x 40 nm rod-shaped virion morphology based on its genome size and genome content. The lack of NAV1 genome similarity to other known viruses suggests that this virus is only distantly related to other archaeal viruses. Previous efforts at characterizing the archaeal virosphere have demonstrated poor connectivity between archaeal viral groups with few exceptions and indicate that most of the viral groups are evolutionary distinct from each other (40). Similar to other Archaeal viruses (41), a majority of the genes encoded by NAV1 are unable to be functionally annotated by sequence analysis alone and most (37/52) have no significant similarity via blastn or blastp in the nr database. While NAV1 is more similar to other archaeal viruses than 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 impacts of viral replication on this 3-way symbiosis.

Acknowledgements

The authors 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 MJY.

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.

The authors declare 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 9:519–25.
3. Markelova NY. 2010. Predacious bacteria, Bdellovibrio with potential for biocontrol. Int J Hyg Environ Health 213:428–431.
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 417:63–7.
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.

6. St. John E, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, Lagutin K, Mitchell K, Reysenbach A-L. 2018. 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.
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.
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.
9. Podar M, Makarova KS, Graham DE, Wolf YI, Koonin E V, 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.
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 5:1–11.
11. Brum JR, Ignacio-espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S. 2015. Patterns and ecological drivers of ocean viral communities. Science (80-) 348:1–10.
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.
13. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang H Bin, 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.
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 E V., 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.
15. Roux S, Adriaenssens EM, Dutilh BE, Koonin E V., 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, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol 37:29–37.

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 5:1–8.
17. John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, Brum J, Polz MF, Boyle E a, Sullivan MB. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environ Microbiol Rep 3:195–202.
18. Nurk S, Meleshko D, Korobeynikov A PP. 2017. metaSPAdes: A New Versatile Metagenomic Assembler. Genome Res 1:30–47.
19. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.
20. Delcher A. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641.
21. Grazziotin AL, Koonin E V, Kristensen DM. 2017. Prokaryotic Virus Orthologous Groups (pVOGs): A resource for comparative genomics and protein family annotation. Nucleic Acids Res 45:D491–D498.
22. 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.
23. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. 2017. Alignment-free d2∗ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res 45:39–53.
24. 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.
25. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.
26. 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.
27. 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.
28. 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.
29. 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 80:7636–44.
30. 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 JVI.02270-15.
31. 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.
32. 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.
33. 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.
34. Quemin ERJ, Quax TEF. 2015. Archaeal viruses at the cell envelope: Entry and egress. Front Microbiol 6:552.
35. 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.
36. 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.
37. 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.
38. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev 40:258–272.
39. 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.
40. Iranzo J, Koonin E V, 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.
41. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin E V. 2018. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. Virus Res 244:181–193.

Figure legends

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

Figure 2. CH09 hot spring pH 2 75˚C where NAV1 was initially sequenced. Electron microscope microbraphs of possible NAV1 morphologies from a CsCl density gradient (insert). White arrows indicate morphologies described in the text. Scale bar is 200nm.

Figure 3. Virus and host relative abundance indicated by read recruitment in three hot spring samples where a cellular metagenome bin was identified as Nanoarchaeota. Relative host (black) and viral (dark grey) abundance in three cellular metagenomes and viral abundance in free viral metaviromes (light grey) from the same hot spring and timepoint.