AEM
Journals.ASM.org

Nanoarchaeota, Their *Sulfolobales* Host, and *Nanoarchaeota* Virus Distribution across Yellowstone National Park Hot Springs

Jacob H. Munson-McGee,[a] Erin K. Field,[b]* Mary Bateson,[c]* Colleen Rooney,[d] Ramunas Stepanauskas,[b] Mark J. Young[c,e]

Department of Microbiology and Immunology, Montana State University, Bozeman, Montana, USA[a]; Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA[b]; Thermal Biology Institute, Montana State University, Bozeman, Montana, USA[c]; Department of Chemistry and Biochemistry, Montana State University, Bozeman, Montana, USA[d]; Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, Montana, USA[e]

***Nanoarchaeota* are obligate symbionts with reduced genomes first described from marine thermal vent environments. Here, both community metagenomics and single-cell analysis revealed the presence of *Nanoarchaeota* in high-temperature (~90°C), acidic (pH ≈ 2.5 to 3.0) hot springs in Yellowstone National Park (YNP) (United States). Single-cell genome analysis of two cells resulted in two nearly identical genomes, with an estimated full length of 650 kbp. Genome comparison showed that these two cells are more closely related to the recently proposed *Nanobsidianus stetteri* from a more neutral YNP hot spring than to the marine *Nanoarchaeum equitans*. Single-cell and catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) analysis of environmental hot spring samples identified the host of the YNP *Nanoarchaeota* as a *Sulfolobales* species known to inhabit the hot springs. Furthermore, we demonstrate that *Nanoarchaeota* are widespread in acidic to near neutral hot springs in YNP. An integrated viral sequence was also found within one *Nanoarchaeota* single-cell genome and further analysis of the purified viral fraction from environmental samples indicates that this is likely a virus replicating within the YNP *Nanoarchaeota*.**

The *Nanoarchaeota* were first described in 2002 (1) as obligate symbionts/parasites with a highly reduced genome lacking most biosynthetic capacity and found in association with their host, the marine hyperthermophilic crenarchaeon *Ignicoccus hospitalis*. The *Nanoarchaeota* provide the unique opportunity to study genome streamlining, as well as the transfer of metabolites between archaeal cells. In their original description, based on 16S ribosomal sequence, *Nanoarchaeum equitans* was classified as a deep-rooted branch of the *Archaea*, forming a new phylum. Although initially accepted, this original classification has been challenged and is still being debated. In part, this debate is based on phylogenetic reconstruction using concatenated protein sequences and analysis of the distribution of gene families among the major archaeal lineages, which suggests a close evolutionary relationship between *N. equitans* and the *Thermococcales*, a basal order of the *Euryarchaeota* phylum (2, 3).

In the 13 years since their discovery, there is evidence that *Nanoarchaeota* inhabit a diversity of environments beyond marine thermal vents. Using primers designed from the 16S rRNA gene of *N. equitans*, genetic evidence of *Nanoarchaeota* has been found to be widespread in terrestrial hot springs (4, 5) and mesophilic hypersaline environments (5). Ribosomal sequences have also been detected from photic-zone water samples far removed from hydrothermal vents (6), suggesting that *Nanoarchaeota* are a widespread and diverse group of *Archaea* capable of inhabiting a broad spectrum of temperatures and geochemical environments. However, despite ribosomal evidence of *Nanoarchaeota* being widespread, the original *N. equitans* isolated with its host *I. hospitalis* remains the only cultured representative of this phylum. No *Ignicoccus* species have been identified in terrestrial hydrothermal samples, suggesting that terrestrial *Nanoarchaeota* use a different host and that *Nanoarchaeota* may associate with a wide diversity of host organisms.

The report of a second sequenced *Nanoarchaeota*, originally termed Nst1 and recently renamed *Nanobsidianus stetteri*, from Yellowstone National Park (YNP), provides more insight into *Nanoarchaeota* diversity (7). Single-cell analysis of *N. stetteri* resulted in a partial 593-kbp genome with an estimated complete genome size of 651 kbp. Analysis of the *N. stetteri* genome provided additional insight into the metabolic capacity and phylogenetic position of the *Nanoarchaeota*. In contrast to *N. equitans*, the full-length genome of *N. stetteri* is estimated to be more than 20% larger (~651 kbp versus 491 kbp), encodes a complete gluconeogenesis pathway, contains no CRISPR system, has components of RNase P, contains a smaller repertoire of split protein coding genes, encodes an euryarchaeal type flagellum (archaellum), and has an inferred *Acidicryptum nanophilum* host.

We investigated the presence of *Nanoarchaeota* species in high-temperature acidic hot springs in YNP and their relatedness to *N. stetteri* and *N. equitans*. We found that *N. stetteri* is present in multiple YNP acidic hot springs, that *Nanobsidianus* species in YNP are found in association with a *Sulfolobales* host, and that a virus is likely replicating within YNP *Nanobsidianus* species.

MATERIALS AND METHODS

Sample collection. Two hot springs were initially selected for this survey (see Fig. S1 in the supplemental material). The location, sampling date,

Received 8 May 2015 **Accepted** 1 September 2015

Accepted manuscript posted online 4 September 2015

Citation 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. doi:10.1128/AEM.01539-15.

Editor: K. E. Wommack

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

* Present address: Mary Bateson, Bioscience Laboratory, Inc., Bozeman, Montana, USA; Erin K. Field, University of Delaware, Newark, Delaware, USA.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01539-15.

pH, and temperature for each sample time are listed in Table S1 in the supplemental material. Hot spring water was collected from the Alice Springs, Crater Hills (CH09), and Nymph Lake 01 (NL01) sampling sites (see Table S1 in the supplemental material). Upon returning to the lab, cells were collected by filtration of samples through 0.4-μm-pore-size filters (Isopore HTTP14250). Cells for catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) analysis were gently washed from filters and fixed in 1% paraformaldehyde for 1 h at room temperature before being washed three times in phosphate-buffered saline (PBS). After fixation, cells were stored in 50% ethanol–50% PBS at −20°C until needed for further processing.

Single-cell genomics. Samples (1 ml) from NL01, collected on 11 October 2011 and 20 September 2012, were preserved with 5% glycerol and 1× Tris-EDTA buffer (final concentrations), immediately frozen in liquid nitrogen, and then stored at −80°C. The flow-cytometric separation of individual cells, cell lysis, whole-genome amplification, and sequencing of 16S rRNA genes were performed at the Bigelow Laboratory Single Cell Genomics Center (scgc.bigelow.org), according to previously described methods (8, 9). Partial 16S rRNA sequences were determined for 99 single cells using universal bacterial and archaeal primers (see Table S2 in the supplemental material). Based on the 16S rRNA results, 13 single cells representing the major archaeal species were selected for whole-genome sequencing. Genomic library preparation and sequencing were performed at the Oregon State University's Center for Genome Research (cgrb.oregonstate.edu). Single-amplified-genome (SAG) genomic DNA was sheared using an S220 focused ultrasonicator (Covaris, Woburn, MA) and gel fractionated for 450-bp fragments. Illumina sequencing libraries were prepared using TruSeq reagents and protocols (Illumina, San Diego, CA). Then, 150×2-bp paired-end reads were sequenced using a HiSeq 2000 platform (Illumina). Twelve indexed SAG libraries were multiplexed, in equal proportions, into one lane of a flow cell. The obtained reads were quality-trimmed, digitally normalized, and assembled in SPAdes v2.2.1 according to the method of Bankevich et al. (10), as described by Wilkins et al. (11). Potential contaminant contigs were removed after tetramer frequency evaluation, and principal component analysis (PCA) identified outliers (12), as well as BLAST comparisons to the NCBI nr database and between samples sequenced together, as described in further detail by Field et al. (13). Genome size estimates were calculated using both arCOG (14) and CheckM (15) analysis.

Hot spring 16S rRNA phylogenetic analysis. Fifty-eight ~510-bp single cell 16S rRNA sequences from NL01 October 2011 and 38 single cell 16S rRNA sequences from NL01 September 2012 were combined with 18 16S rRNA sequences from NL01 metagenomes (16, 17) and 21 sequenced reference genomes, resulting in a total of 135 sequences. These sequences were aligned using MUSCLE (18), and a Bayesian analysis was performed using MrBayes (version 3.2.5) (19), with mixed nucleotide substation models and a gamma-shaped parameter. Posterior probability values were derived from 10 million permutations, while sampling every 100,000 generations using the default 25% burn-in. The 16S rRNA sequence from the bacterial strain *Acidithiobacillus caldus* ATCC 51756 served as the outgroup.

CARD-FISH probe design. A total of eight 16S rRNA probes were designed, each of which detects one of the eight major 16S rRNA phylotypes in the NL01 hot spring (see Table S2 in the supplemental material). An additional 16S rRNA probe was used to detect most *Crenarchaeaota* (20). Probes were synthesized by Biomers (Ulm, Germany) with horseradish peroxidase incorporated at their 5′ ends. The specificities of the probes were validated by testing them in the lab against nearly full-length 16S rRNA clones generated for each major 16S rRNA present in NL10. Probes were confirmed to hybridize to only their intended target and not to the other 16S rRNA types present in NL10.

CARD-FISH analysis. A modified CARD-FISH analysis (21, 22) was used to probe *Archaea*-dominated acidic hot spring environmental samples. Fixed samples were placed in the wells of glass slides (PL-2026 Precision Lab Products) and air dried for 10 min at 46°C. Samples were subsequently dehydrated in 50, 80, and 100% ethanol for 3 min and dried at 46°C for 5 min. The wells were covered with permeabilization solution (50 mM glucose, 20 mM Tris [pH 7.5], 10 mM EDTA, and 0.2% Tween 20) and placed on ice for 1 h before being rinsed in 1× PBS and air dried at 46°C. Endogenous peroxidases were deactivated by a 10-min incubation in 0.1 N HCl at room temperature, rinsed with 1× PBS, and air dried at 46°C. 16S rRNA probes were added to the hybridization buffer to a final concentration of 0.2 ng/μl with various amounts of formamide (*Sulfolobales*, 20%; *Nanobsidianus*, 40%) to increase the stringency of the hybridization. Samples were allowed to hybridize at 46°C in petri plates sealed with parafilm for 3 h before being washed in washing buffer (22) at 48°C for 20 min. This was followed by 15 min of rinsing in 1× PBS at 37°C before air drying at 46°C. All samples were overlaid with a solution containing 1× PBS, 10% dextran sulfate, 0.1% blocking reagent (Roche [Germany], catalog no. 11096176001), 2 M NaCl, 0.0015% H_2O_2, and 0.33 μg of Alexa Fluor 488- or Alexa Fluor 594-labeled tyramides/μl and then incubated at 37°C for 30 min. Washes consisted of 5 min with 1× PBS at 46°C and 1 min with water, followed by air drying. For dual labeling, the protocol was repeated starting at the deactivation step using different 16S ribosomal probes and tyramide fluorophore. After CARD amplification, slides were washed in 1× PBS, stained with DAPI (4′,6′-diamidino-2-phenylindole) at 10 ng/μl and fixed with Vectashield. Controls of equivalent *Escherichia coli* rRNA probes (pB-02228), competition with nonfluorophore-labeled probes, and nuclease treatments were also performed.

Samples were imaged on a Leica TBS SP8 confocal microscope fitted with a 63× oil immersion lens, and images were collected sequentially. NIH ImageJ 64 software was used to process the images.

Geographic distribution of *Nanobsidianus* within YNP. Seven conserved genes (see Table S3 in the supplemental material) contained within two single-cell genomes described here (AB_777_F03 [F03] and AB_777_O03 [O03]), *N. stetteri*, *N. equitans*, and CH09 metagenomes (17) were concatenated and aligned using MUSCLE (18) and subjected to Bayesian analysis with MrBayes (18) as described above. In addition, we recruited contigs from 23 publicly available metagenomes (17) using the O03 contigs. A minimum overlap of 50 bp and a 93% nucleic acid identity were used to recruit contigs onto the O03 genome.

***Nanobsidianus* virus identification.** The single-cell genomes were analyzed with the VirSorter app (https://de.iplantcollaborative.org/de/) to identify viral sequences in the single-cell genomes. After manual curation of the sequences, identified genes were subjected to BLASTx analysis to search for matches to viral proteins with a minimum E value of 10E–5. Identified regions were additionally subjected to BLAST analysis against viral network clusters described by B. Bolduc et al. (40).

To gain further evidence for the presence of a *Nanobsidianus*-associated virus, viral fractions were prepared by filtering NL01 hot spring water through a 0.45-μm-pore-size filter to remove cells. Viruses in the filtered water were concentrated by $FeCl_3$ flocculation (23). Virus particles were separated by buoyant cesium chloride density centrifugation gradients spanning 1.13 to 1.38 g/ml (116,000 × *g* for 20 h). The total gradient was hand fractionated and DNA was extracted from individual fractions using a PureLink viral RNA/DNA minikit (Invitrogen, Waltham, MA). Fractions were screened for the presence of the viral genome by PCR analysis using virus-specific genome primers. PCR primers specific to *Nanobsidianus* and *Sulfolobales* were included as a control (see Table S2 in the supplemental material). Total viral fractions from CH09 were also prepared by filtration to remove cells and $FeCl_3$ flocculation.

RESULTS

Ninety-six 16S rRNA sequences derived from single cell sequencing and 18 16S rRNA sequences from cellular metagenomes both showed a simple archaeal microbial community present in NL01 (see Fig. S2 in the supplemental material). This community is dominated by nine species based on their 16S rRNA genes. Based on metagenomic read abundance and single cell sequencing, eight

archaeal species represent 97% of all cells, and a single bacterial *Hydrogenobaculum* sp. comprises the remaining 3%. The eight archaeal clades are either related to *Acidianus hospitalis* (3%) and *Sulfolobus islandicus* (1%) or uncultured members of what likely constitute new crenarchaeal species, the recently proposed *Acidicryptum nanophilium* (46%), an *Acidolobus* sp. (6%), a *Vulcanisieta* sp. (5%), a *Nanobsidianus* sp. (16%), and two *Sulfolobus* spp. accounting for 11 and 9%, respectively (see Fig. S2 in the supplemental material).

We selected representative single cells from each of the eight archaeal clades for genome sequence analysis based on their 16S rRNA sequence (see Table S4 in the supplemental material). We were successful in obtaining high quality sequence information for 13 single cells. The estimated genome completeness for sequenced single cells ranged from 13 to 85%. The DNA sequence and assembly statistics are provided (see Table S5 in the supplemental material).

Two of the thirteen single-cell genomes represented within NL01, designated O03 and F03, were classified as *Nanobsidianus* based on their 16S rRNA sequences. These two SAGs had high 16S rRNA gene homology (98.2%) to *N. stetteri* sequenced from Obsidian Pool in YNP and low homology to *N. equitans* (81.5%), which was isolated from a marine hydrothermal vent. A blastn analysis of four metagenomes from the same hot spring sampled over a 7-year period (minimum E score of 1E–5) with the partial O03 16S rRNA gene revealed *Nanobsidianus* 16S rRNA gene sequences in all four metagenomes.

An examination of multiple YNP hot springs revealed that *Nanobsidianus* is broadly distributed. A BLASTn search of metagenomes from 23 hot springs (minimum E value of 1E–5) with the O03 single-cell partial *Nanobsidianus* genome (described below) resulted in 7,024 contigs from 15 of the 23 hot springs (see Table S6 in the supplemental material). Ten of the fifteen hot springs also had *Nanobsidianus* 16S rRNA sequences (see Table S7 and Fig. S3 in the supplemental material). Based on the read abundance, *Nanobsidianus* cells were most abundant in hot springs that had temperatures of >60°C and a pH of <4.5 (Fig. 1A). Outside this temperature and pH range, *Nanobsidianus* sequences were rarely detected. Read recruitment analysis indicated the *Nanobsidianus* represents ca. 11% of the NL01 and 5% of the CH09 microbial community composition (Fig. 1B). However, one should be cautious in assigning quantitative community membership based on single cell genomics and metagenomic read recruitment analysis due to the inherent bias that each technique presents (24–28).

Assembly of the two *Nanobsidianus* genomes derived from two independent single cells produced two incomplete genomes of 499 kbp (F03) and 549 kbp (O03). However, upon further analysis, 50 kbp of the F03 genome was revealed to be from a *Sulfolobales* species (described below), resulting in a genome of 449 kbp, which was used in further analysis. The degree of genome completeness was estimated using conserved archaeal genes that are represented in the assembled genomic contigs (14). In the arCOG database as of March 2015, there were 95 genes conserved across all 168 sequenced archaeal genomes, including *N. equitans*, and 325 genes in all *Crenarchaeota* genomes. The completeness of *Nanoarchaeota* genomes was calculated using the 95 genes conserved in all archaeal genomes, while the genome completeness of all other SAGs was calculated using the 325 genes that are shared by all *Crenarchaeota*. F03 contained 74 of 95 genes, resulting in an esti-

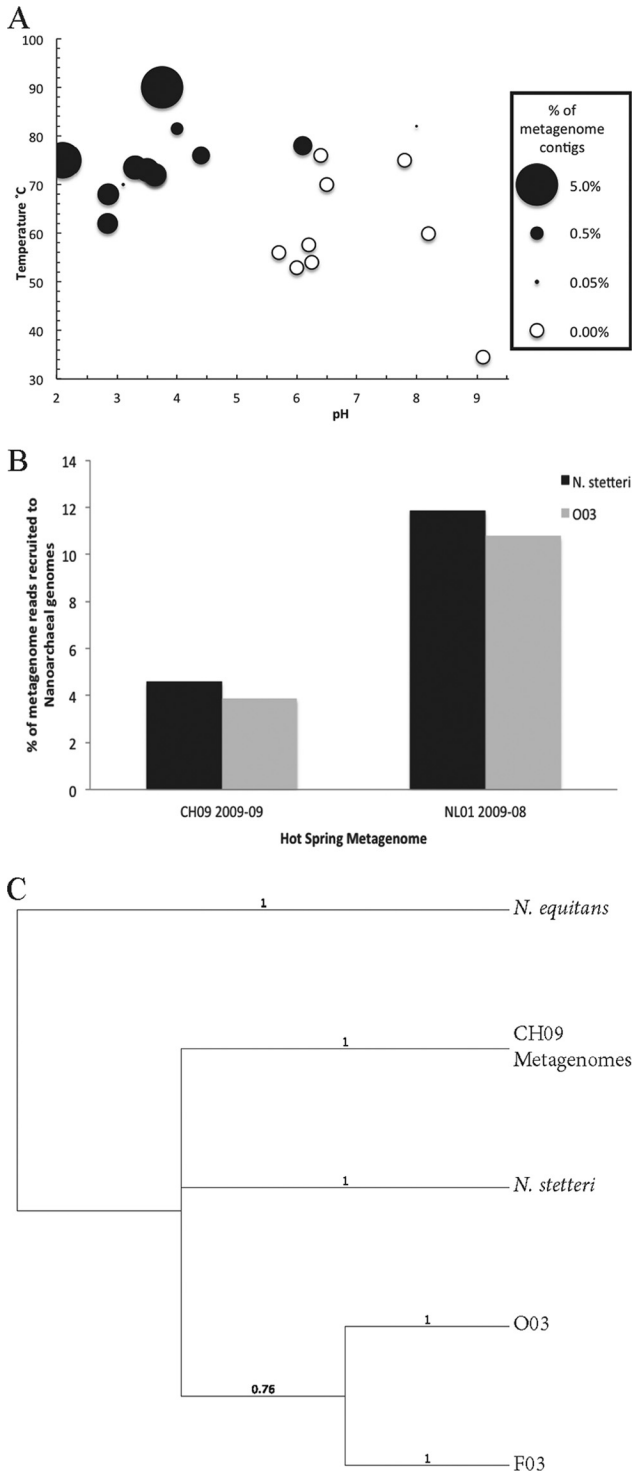

FIG 1 (A) Distribution of *Nanoarchaea* in 22 hot springs in Yellowstone National Park based on hot spring pH and temperature. The sizes of the solid black circles indicate the percentage of the metagenome contigs that were recruited to the *N. stetteri* single-cell genome O03, while open circles indicate hot spring metagenomes with no nanoarchaeal sequences recruited. (B) Recruitment of metagenomic reads to *N. stetteri* Obsidian Pool (gray columns) and *N. stetteri* O03 (black columns) partial genomes, indicating the relative abundances of *Nanobsidianus* in two YNP hot springs. (C) Transformed maximum-likelihood phylogenetic tree of seven concatenated conserved *Nanoarchaeota* genes in F03, O03, *N. stetteri*, *N. equitans*, and CH09 cellular metagenomes. Numbers indicate the posterior probability and *N. equitans* served as the outgroup.

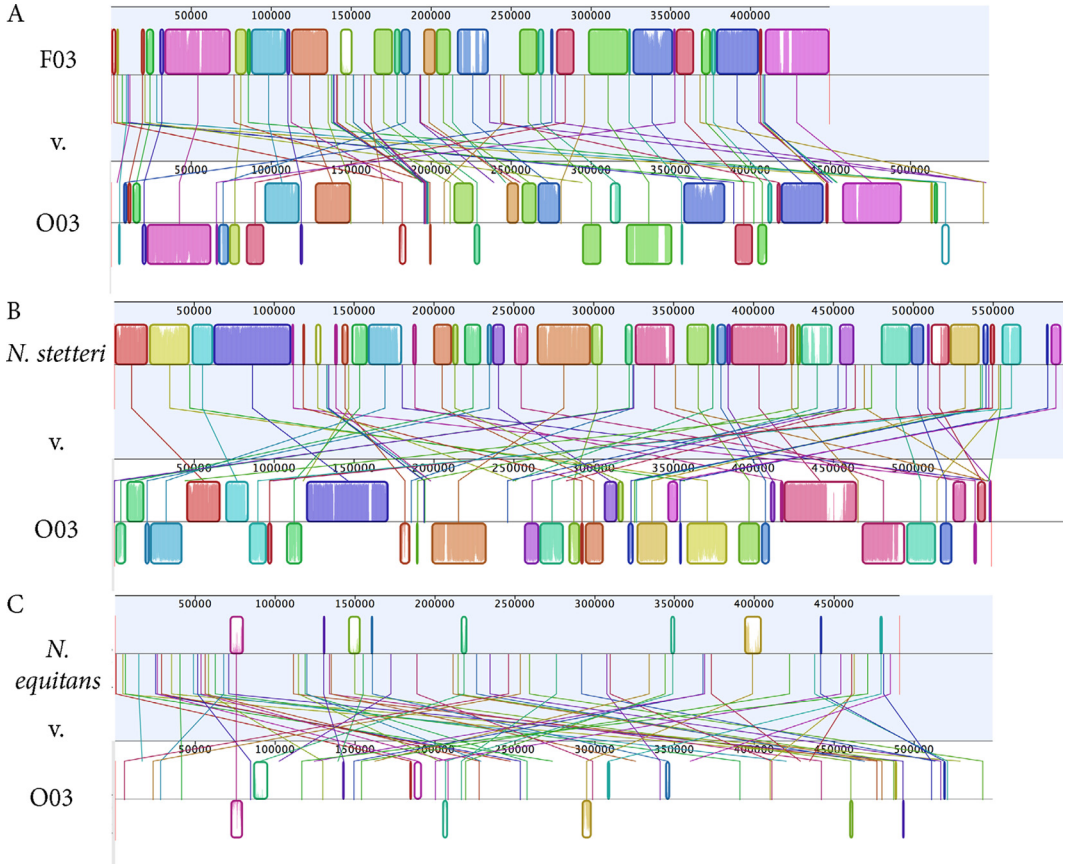

FIG 2 Mauve alignment of O03 concatenated partial length genome to F03 (A), *N. stetteri* (B), and full-length *N. equitans* (C) genomes. Blocks of the same color indicate regions of homology between the genomes, and blocks below the genome's center line are inverted relative to the reference sequence.

mated genome size of 576 kbp with 95% confidence limits of 530 to 644 kbp; O03 has 76 of 95 conserved genes and an estimated genome size of 686 kbp and 95% confidence limits of 635 to 762 kbp (see Table S5 in the supplemental material). The discrepancies in genome length between these two cells could reflect an estimation bias resulting from the uneven distribution of the conserved genes used in the analysis, a true difference in their genome size, and/or the potential presence of sequences from the *Nanobsidianus* host (discussed below). Based on the calculated genome size of *N. stetteri* (650 kbp), F03 is 67% complete, and O03 is 85% complete.

The partial genome sequences of *Nanobsidianus* from NL01 show a high degree of sequence identity to each other (Fig. 2; see also Fig. S4 in the supplemental material) (29). Of the 472 open reading frames (ORFs) that were identified in both F03 and O03, 372 (79%) encoded products with >95% amino acid identity to each other. Of 532 ORFs identified in both O03 and *N. stetteri*, 309 (58%) encoded products with >95% amino acid identity to each other. In contrast, none of the 313 ORFs identified in both O03 and *N. equitans* encoded products with ≥95% amino acid identity. Whole-genome Mauve alignment showed more conserved regions between the three YNP *Nanobsidianus* genomes compared to *N. equitans* (Fig. 2) (29). The construction of a maximum likelihood tree with seven concatenated conserved genes in F03, O03, *N. stetteri*, *N. equitans*, and *Nanobsidianus* contigs identified in CH09 metagenomes supports the close relationship between YNP *Nanobsidianus* species compared to the marine *N. equitans* (Fig. 1C). The terrestrial *Nanobsidianus* sequences from YNP grouped together, while *N. equitans* formed its own distinct branch. Overall, these results support the conclusion that the F03 and O03 genomes are highly homologous and more closely related to *N. stetteri* than to *N. equitans*.

Additional analysis of specific genes supports the overall conclusion that F03 and O03 are closely related to *N. stetteri*, but there are some differences. In O03 we identified Rpp29 and Rpr2, two components of RNase P. We also identified a truncated version (238 bp) of the RNase P RNA component in both of the genomes that is 98% similar to the one described by Podar et al. (7). The presence of these genes is similar to *N. stetteri*, in which three protein components were identified, but is different from *N. equitans*, which remains the only identified cellular life that lacks this complex (30). We also searched for the twelve split protein coding genes previously described in *N. equitans* and *N. stetteri* (7). All but the P-loop ATPase were found in at least one of the two partial NL01 *Nanobsidianus* genomes. All YNP single cells had identical split protein coding genes. Of these split genes, six are split in the same place as *N. equitans*; four genes that are split in *N. equitans* are not split in YNP single cells, and two are split in YNP isolates but not in *N. equitans* (see Table S8 in the supplemental material).

We searched the NL01 *Nanobsidianus* genomes for tRNA genes using a tRNA scan (31). Between the two NL01 single cell genomes, all 20 standard amino acids have at least one tRNA repre-

sented. We did not find any nonstandard tRNAs in either of these genomes. Within the two genomes, we found 36 total tRNA sequences (see Table S9 in the supplemental material). In contrast, *N. stetteri* is missing a tRNA for Phe, most likely due to the incompleteness of the genome (7, 32). Both F03 and O03 have four intron-containing tRNA genes, in contrast to *N. stetteri*, which only has two intron containing tRNAs (see Table S9 in the supplemental material). This discrepancy could again be explained by the incompleteness of the *N. stetteri* genome. Of the four intron-containing tRNA genes in F03 and O03, two—Ile (TAT) and Tyr (GTA)—are also found in *N. stetteri* and *N. equitans*. One, Met (CAT), is not split in *N. stetteri* and has a longer intron (65 versus 27 bp) in *N. equitans*, and Leu (TAA) is not split in *N. equitans* and is not detected in *N. stetteri*. In addition, like *N. equitans* and *N. stetteri*, F03 and O03 lacked central metabolic genes, including nearly all of the genes for lipid, cofactor, amino acid, and nucleotide synthesis, indicating a symbiotic lifestyle where the *Nanobsidianus* cells are incapable of replicating without their host (33). All of the archaeal flagellum genes contained in *N. stetteri* described by Podar et al. (7) are contained in at least one of the NL01 *Nanobsidianus* genomes except for a FlaH homolog (the putative ATPase) gene. The NL01 *Nanobsidianus* genomes also lacked any genes associated with the CRISPR/Cas system; this finding is similar to that for *N. stetteri* but in contrast to that for *N. equitans*, in which crRNAs appear to be constituently expressed (34).

Analysis of the assembled F03 sequences revealed the presence of a *Sulfolobales* organism that may serve as a host for the nanoarchaeal cell. Of the 25 assembled contigs from F03 (representing 50 kbp of the DNA sequence or 9.9% of the total assembled sequences), 5 did not map to any known *Nanoarchaeota* genomes. These 5 contigs also have a significant G/C skew (46% G/C) compared to the 20 contigs that have homology to the O03 and *N. stetteri* genomes (24%). A BLAST analysis of these five contigs against the 11 other SAGs reveals that each contig has the highest match to SAGs AB_777_J03, AB_777_K09, or AB_777_K20 (here referred to as J03, K09, and K20, respectively), all members of the dominant *Sulfolobales* group, identified as *A. nanophilum* by Podar et al. (7). In addition, one of these contigs contained a CRISPR locus with a conserved leader and direct repeat sequences frequently found in *Sulfolobus* species and matched those found in the single-cell genomes of J03, K09, and K20. These contigs are likely present as a result of an intimate association between the *Nanobsidianus* and *Acidicryptum* cells, namely, that of host and symbiont. One single-cell 16S rRNA PCR amplicon that was not selected for genome sequencing had 16S rRNA signatures from two organisms. One of the 16S rRNA sequences was from *A. nanophilum* and was 99.1% similar to one from J03, while the second resembles one from *Nanobsidianus* and was 99.1% similar to one from O03. Additionally, *A. nanophilum* 16S rRNA signatures are present in all of the YNP metagenomes where *Nanobsidianus* 16S rRNA sequences are found. All of these lines of evidence suggest that the *Nanobsidianus* forms a host-symbiont relationship with *A. nanophilum* present in the NL01 hot springs.

To further investigate the nature of the host, CARD-FISH analysis was performed to confirm the genomic host identification of the *Nanobsidianus* present in the NL01 and CH09 hot springs (Fig. 3). Briefly outlined, fluorescently labeled hybridization probes specific to the each of the eight major cell types were generated (see Materials and Methods). The probe specific to *Nanobsidianus* rRNA was labeled with Alexa Fluor 594 fluorophore, while the rRNA probes to the other seven individual cell types in NL01 were labeled with Alexa Fluor 488. The *Nanobsidianus* probe was used in combination with each individual cellular rRNA probes in a dual hybridization assay of cells collected from NL01 and CH09. CARD-FISH analysis revealed that NL01 and CH09 *Nanobsidianus* strains are only found in association with an *Acidicryptum* species (Fig. 3A, B, C, D, and G) and not in association with the other six archaeal species present in NL01 (Fig. 3E). This confirms the host species identified by single-cell genomics analysis. CARD-FISH analysis demonstrated colocalization of >95% of detected *Nanobsidianus* cells with *Acidicryptum* cells in both NL01 and CH09 hot springs. Further analysis showed the same colocalization in hot spring samples collected over the course of 17 months (Fig. 3F), indicating that the *Nanobsidianus* population is a persistent member of YNP hot spring environments. Controls of *Nanobsidianus* cells dually labeled with the other six potential host species in the hot springs showed no colocalization, indicating that the host-symbiont relationship is limited to this *Acidicryptum* host.

Analysis of colocalization of the *Acidicryptum* host and the *Nanobsidianus* cells revealed fluctuations of both over time in both hot springs surveyed. In CH09 the *Acidicryptum* varied from 7 to 23% of the total microbial community composition, and 53 to 75% of *Acidicryptum* cells were associated with *Nanobsidianus*. In NL01 the proportion of the *Acidicryptum* cells ranged from 17 to 45% of the total microbial community. Regardless of their relative abundance, ~66% of these cells were associated with *Nanobsidianus* cells (see Table S10 in the supplemental material). Overall, these results indicate that the *Acidicryptum* is the host for the *Nanobsidianus* in NL01 and CH09 hot springs.

***Acidicryptum* host genome analysis.** The three *Acidicryptum* host single-cell genomes, J03, K09, and K20, showed high sequence identity to *A. nanophilum*, the host of *N. stetteri* proposed by Podar et al. (7). These three partial genomes, all from of the *Sulfolobales* order, range from 250 to 850 kbp. Of the 325 conserved genes in all *Crenarchaeota* genomes, J03 contains 260, K09 663, and K20 157, yielding estimated genome size and genome completeness values using the arCOG database of 1.070 Mbp (80.0%), 1.034 Mbp (20.3%), and 1.078 Mbp (48.3%), respectively (see Table S5 in the supplemental material). Genome size estimates were independently analyzed with CheckM (15), which estimated an average genome size of 1.149 Mbp for the three *N. stetteri* host cells. This estimate of ~1.1 Mbp is significantly smaller than the estimated size of the *A. nanophilum* genome (1.7 Mbp) from Obsidian Pool previously described (7). It is possible that these methods are underestimating the host genome size due to the inherent uncertainty of using partial genome sequences combined with the unevenness of multiple displacement amplification methods. *I. hospitalis*, the host of *N. equitans*, has a genome size of 1.29 Mbp and has also undergone genome streamlining (7).

It has been proposed that *Nanoarchaeota* evolve faster then nonsymbiotic *Archaea*, as a consequence of their reduced genome size and content (3). We observe this trend with *Nanobsidianus* cells from the different YNP hot springs showing more divergence from each other compared to the YNP *Acidicryptum*. Of ORFs identified in both J03 and *A. nanophilum*, 79% encode products with >95% amino acid identity (see Fig. S5 in the supplemental material), whereas just 58% of the ORFs in O03 and *N. stetteri* encode products with >95% amino acid identity.

***Nanobsidianus* virus detection.** Analysis of the *Nanobsidianus*

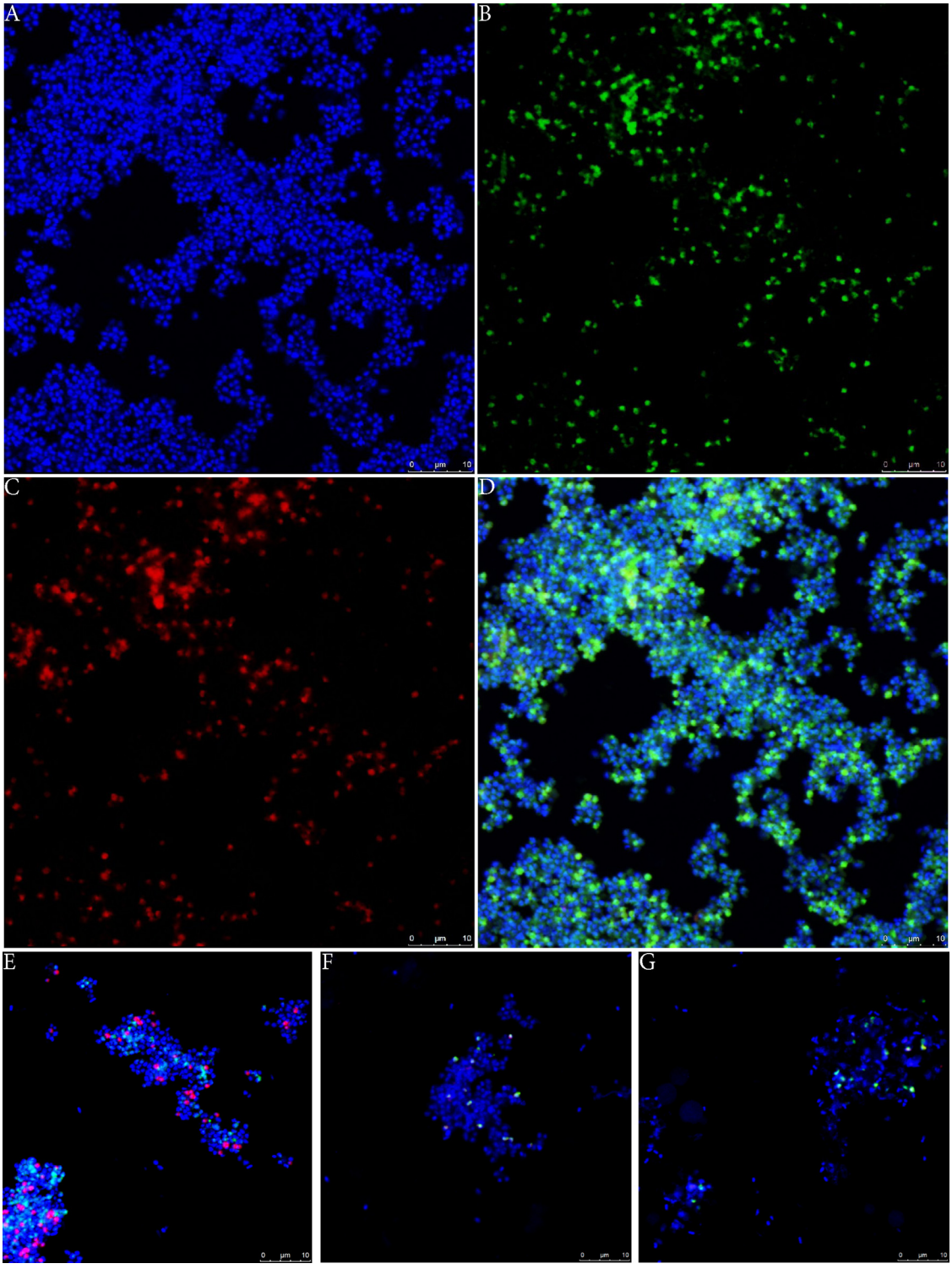

FIG 3 CARD-FISH analysis of *N. stetteri*-host interaction from environmental samples. (A to C) Total cells from CH09 2012-01 stained blue with DAPI (A) and cells imaged with 16S ribosomal probe to *Acidicryptum* host cells (green, B). Cells imaged with 16S ribosomal probe to *Nanobsidianus* are also shown (red, C). (D) Merged image of panels A to C indicating colocalization of *Acidicryptum* and *N. stetteri* cells. (E) Merged control image of *Acidolobus* cells (green) and *Nanobsidianus* cells (red) from the same hot spring and sampling date showing the lack of colocalization. (F and G) Additional merged images of *Nanobsidianus* cells (red) and *Acidicryptum* host cells (green) from a different sampling date CH09 2013-05 (F) and hot spring NL01 2014-09 (G).

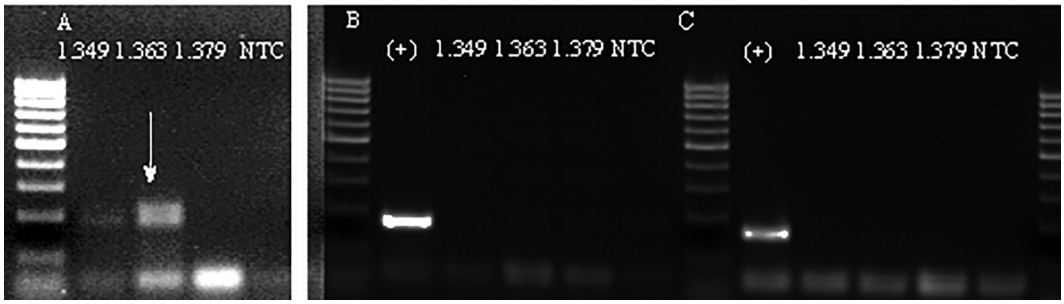

FIG 4 PCR analysis of virus purified from NL01 environmental sample. Fractions of CsCl buoyant density gradients were tested for a viral genome segment (A) and for contamination of *Acidicryptum* (B) or *N. stetteri* (C) cellular sequences in the same viral fractions. Numbers indicate the density of the CsCl gradient in g/cm^3, "(+)" indicates a positive-control template, and NTC indicates the no-template control. An arrow indicates the presence of the expected viral product.

O03 single cell genome with VirSorter identified one 10-kbp contig as having a virally enriched region. Annotation of this contig showed 10 ORFs, 5 of which were annotated as encoding hypothetical unknown proteins. Interestingly, two other ORFs were annotated, one as encoding a hypothetical protein related to a *Sulfolobus* monocaudavirus 1 (SMV1) conserved archaeal viral protein and the other as encoding an *N. equitans* 30S ribosomal protein, suggesting that this contig contains a portion of an integrated viral genome. The archaeal virus-related ORF is located at the end of the 10-kbp contig and has a GC content of 29.8%, which is significantly different from the rest of the contig on which it is located (22.6%) and the GC content of all O03 contigs (23.4%). Further BLAST analysis of this contig against a collection of viral metagenomes from NL01 resulted in 104 contigs matching to the conserved archaeal viral protein. Furthermore, all of these 104 contigs mapped to a single viral network cluster of the viral community structure previously reported (40). Assembly of this network cluster generated a 3.0-kbp contig with the SMV1 conserved archaeal virus-like protein coding sequence at one end and an ~600-bp putative structural protein coding sequence from a *Sulfolobus* rudivirus near the other end (see Fig. S6 in the supplemental material), as well as several other ORFs with no significant similarity. The GC content of the identified conserved archaeal virus gene (29.8%) is nearly identical to the GC content of all of the contigs that make up the viral network cluster (29.9%). The presence of this ORF in the viral fraction of NL01 hot spring water was confirmed by PCR analysis of virus-like particles isolated from CsCl buoyant density gradients. This genome segment was successfully PCR amplified and confirmed by DNA sequencing from the purified viral fraction on cesium chloride fractions with a density of 1.36 g/cm^3. The same genome segment was also successfully amplified from total viral fractions from CH09, suggesting that the viral distribution is potentially similar to *Nanobsidianus* in YNP. In order to show that cells were not contaminating the viral fraction, PCR with primers specific to both *Nanobsidianus* and *Acidicryptum* spp. was performed on the same fractions, but no sequence was PCR amplified (Fig. 4). Overall, these results indicate that there is likely a virus associated with *Nanobsidianus* present in YNP hot springs.

DISCUSSION

We report here the sequencing of two *Nanobsidianus* single-cell genomes from high temperature acidic hot springs in YNP, and the detection of *Nanobsidianus* cells from high-temperature acidic hot springs across YNP. Based on their high 16S rRNA similarity (>95%) (35) to *N. stetteri*, and their overall genome homology, we propose that these cells are *N. stetteri* and that they share the same *A. nanophilum* host as previously described (7). All three YNP *Nanobsidianus* genomes are closely related to each other and quite distinct from *N. equitans*. Even though members of the *Nanobsidianus* genus are widely distributed in YNP thermal features, including in Yellowstone Lake (6), they are most abundant in high temperature acidic hot springs found in YNP.

N. stetteri cells found in the Obsidian Pool, NL01, and CH09 all share an *A. nanophilum* host. In our analysis, ~66% of *A. nanophilum* cells are found in association with *N. stetteri* cells. The reason why this approximate ratio in environmental samples is maintained within different hot springs is unknown, but this observation suggests that there is some control over these cell interactions. It remains to be determined whether the physical interaction between the YNP *N. stetteri* and its host is similar to *N. equitans* and its *Ignicoccus hospitalis* host. Our estimates of a 1.1-Mbp genome for the *A. nanophilum* host suggests an expanded mutualism with its *Nanobsidianus* partner in the NL01 hot spring system. However, the ability to find *A. nanophilum* lacking its *Nanobsidianus* partner indicates that *A. nanophilum* is capable of independent replication, making it the smallest genome of a free-living organism. This speculation will need to be confirmed by complete genome sequencing and culturing of *A. nanophilum* from *NL01*. The fact that the marine *N. equitans* and terrestrial YNP *N. stetteri* are associated with different hosts suggests additional *Nanoarchaeota*-host interactions will be discovered in other environments and that there is a high degree of flexibility in the host-*Nanoarchaeota* partnership found in nature. It is clear that independent *Nanoarchaeota*-host partnerships have arisen over time, and it also suggests that the *Nanoarchaeota*-host association may be an old and essential feature of the archaeal communities.

Reno et al. (36) showed that microorganisms in the same geographic region are more closely related to each other than members of the same species isolated from geographically distinct areas. Although that study was limited to major geographical areas, its results likely can be applied to the phylogenetic distribution of *N. stetteri* in YNP. Cells from the same hot spring are more closely related to each other than to cells from different hot springs, suggesting that *N. stetteri* in each hot spring may be independently adapting to local hot spring environments.

Similar to previous descriptions of *N. stetteri*, we did not observe a CRISPR/Cas system, which is in contrast to *N. equitans*, where the CRISPR/Cas system was detected (37). Although it is

possible that the three partial genomes of *N. stetteri* from YNP have all missed the CRISPR/Cas system, we find this unlikely. It is noteworthy that all *Nanoarchaeota* genomes contain split tRNA genes. The absence of the CRISPR/Cas system in some *Nanoarchaeota* that contain split tRNAs diminishes the argument that split tRNAs would be unlikely to arise in *Nanoarchaeota* due to the ability of the CRISPR/Cas system to eliminate foreign genetic elements (38). We also find it more likely that the marine lineage gained the CRISPR/Cas system as opposed to the terrestrial lineage losing this system, since the YNP genomes are significantly larger. The likely presence of a virus within *N. stetteri* diminishes the argument that the lack of viral pressure is responsible for the loss of the CRISPR/Cas system and provides support for split tRNAs arising as a mechanism to escape integration of mobile genetic elements in the anticodon loop (39).

Our results are consistent with those of Podar et al. (7) but also suggest a widespread distribution of *Nanoarchaeota* in the thermal features of YNP. This expanded collection of sequences provides a foundation for future studies on *Nanoarchaeota* genome reduction. The current model of *Nanoarchaeota* presumes that symbiont and host have evolved together to the point where *N. equitans* is only able to grow with *I. hospitalis* and not with any closely related *Ignicoccus* sp. (33). It remains to be seen whether the *Nanobsidianus* spp. of YNP share the host limitations of *N. equitans* or if *Nanobsidianus* spp. from one hot spring are able to grow with a host species isolated from a different hot spring.

The detection of a virus likely replicating within *Nanobsidianus* cells is the first report of a virus associated with ultrasmall *Archaea* (*Nanoarchaeota*, *Parvachaeota*, and *Nanohaloarchaeota*). The discovery of such a virus raises interesting questions about the possible three-way interactions between the virus and *Nanobsidianus* and *Acidicryptum* cells. To our knowledge, this is the smallest host genome supporting virus replication. This highly reduced host genome, which contains all of the tRNA genes but lacks most of the genes required for central carbon metabolism, provides an opportunity to examine the minimum requirements for viral replication.

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation grants DEB-4W4596 (to M.J.Y.) and DEB-1441717 and DBI-1226726 (to R.S.).

We thank Jennifer Wirth and Ross Hartman for critical reading of this text. We also thank Matthew Lavin for advice regarding phylogenetic analysis. The research was conducted in Yellowstone National Park under the conditions of permits YELL-2011-SCI-5090, YELL-2012-SCI-5090, and YELL-2013-SCI-5090. We also thank the YNP research resource office and especially Stacey Gunther for their work facilitating sampling within YNP.

We declare no conflict of interest.

REFERENCES

1. **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. http://dx.doi.org/10.1038/417063a.
2. **Petitjean C, Deschamps P, López-García P.** 2014. Rooting the domain *Archaea* by phylogenomic analysis supports the foundation of the new kingdom *Proteoarchaeota*. Genome Biol Evol **7:**191–204. http://dx.doi.org/10.1093/gbe/evu274.
3. **Brochier C, Gribaldo S, Zivanovic Y, Confalonieri F, Forterre P.** 2005. *Nanoarchaea*: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to *Thermococcales*? Genome Biol **6:**R42. http://dx.doi.org/10.1186/gb-2005-6-5-r42.
4. **Hohn MJ, Hedlund BP, Huber H.** 2002. Detection of 16S rRNA sequences representing the novel phylum "*Nanoarchaeota*": indication for a wide distribution in high temperature biotopes. Syst Appl Microbiol **25:**551–554. http://dx.doi.org/10.1078/07232020260517698.
5. **Casanueva A, Galada N, Baker GC, Grant WD, Heaphy S, Jones B, Yanhe M, Ventosa A, Blamey J, Cowan DA.** 2008. Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. Extremophiles **12:**651–656. http://dx.doi.org/10.1007/s00792-008-0170-x.
6. **Clingenpeel S, Kan J, Macur RE, Woyke T, Lovalvo D, Varley J, Inskeep WP, Nealson K, McDermott TR.** 2013. Yellowstone lake nanoarchaeota. Front Microbiol **4:**274. http://dx.doi.org/10.3389/fmicb.2013.00274.
7. **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. http://dx.doi.org/10.1186/1745-6150-8-9.
8. **Stepanauskas R, Sieracki ME.** 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci U S A **104:**9052–9057. http://dx.doi.org/10.1073/pnas.0700496104.
9. **Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, Reinthaler T, Poulton NJ, Masland EDP, Gomez ML, Sieracki ME, DeLong EF, Herndl GJ, Stepanauskas R.** 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science **333:**1296–1300. http://dx.doi.org/10.1126/science.1203690.
10. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
11. **Wilkins MJ, Kennedy DW, Castelle CJ, Field EK, Stepanauskas R, Fredrickson JK, Konopka AE.** 2014. Single-cell genomics reveals metabolic strategies for microbial growth and survival in an oligotrophic aquifer. Microbiology **160:**362–372. http://dx.doi.org/10.1099/mic.0.073965-0.
12. **Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng J-F, Eisen JA, Sieracki ME, Stepanauskas R.** 2009. Assembling the marine metagenome, one cell at a time. PLoS One **4:**e5299. http://dx.doi.org/10.1371/journal.pone.0005299.
13. **Field EK, Sczyrba A, Lyman AE, Harris CC, Woyke T, Stepanauskas R, Emerson D.** 2014. Genomic insights into the uncultivated marine zeta-proteobacteria at Loihi Seamount. ISME J **9:**857–870. http://dx.doi.org/10.1038/ismej.2014.183.
14. **Wolf YI, Makarova KS, Yutin N, Koonin EV.** 2012. Updated clusters of orthologous genes for *Archaea*: a complex ancestor of the *Archaea* and the byways of horizontal gene transfer. Biol Direct **7:**46. http://dx.doi.org/10.1186/1745-6150-7-46.
15. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res **25:**1043–1055. http://dx.doi.org/10.1101/gr.186072.114.
16. **Snyder JC, Bateson MM, Lavin M, Young MJ.** 2010. Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. Appl Environ Microbiol **76:**7251–7258. http://dx.doi.org/10.1128/AEM.01109-10.
17. **Inskeep WP, Jay ZJ, Tringe SG, Herrgård MJ, Rusch DB.** 2013. The YNP Metagenome Project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol **4:**67. http://dx.doi.org/10.3389/fmicb.2013.00067.
18. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:**113. http://dx.doi.org/10.1186/1471-2105-5-113.
19. **Huelsenbeck JP, Ronquist F.** 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755. http://dx.doi.org/10.1093/bioinformatics/17.8.754.
20. **Loy A, Maixner F, Wagner M, Horn M.** 2007. probeBase—an online resource for rRNA-targeted oligonucleotide probes: new features 2007. Nucleic Acids Res **35:**D800–D804. http://dx.doi.org/10.1093/nar/gkl856.
21. **Brileya KA, Camilleri LB, Fields MW.** 2014. 3D-fluorescence in situ hybridization of intact, anaerobic biofilm, p 189–197. *In* Sun L, Shou W

(ed), Engineering and analyzing multicellular systems. Springer, New York, NY.
22. **Pernthaler A, Pernthaler J.** 2004. Sensitive multicolor fluorescence in situ hybridization for the identification of environmental microorganisms. Mol Microb Ecol Man **3:**711–726.
23. **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. http://dx.doi.org/10.1111/j.1758-2229.2010.00208.x.
24. **Lasken RS, Stockwell TB.** 2007. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnol **7:**19. http://dx.doi.org/10.1186/1472-6750-7-19.
25. **Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Hwai W, Watson D, Brodie EL, Hazen TC, Chang HW, Keller M.** 2006. Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl Environ Microbiol **72:**3291–3301. http://dx.doi.org/10.1128/AEM.72.5.3291-3301.2006.
26. **Yilmaz S, Allgaier M, Hugenholtz P.** 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat Methods **7:**943–944. http://dx.doi.org/10.1038/nmeth1210-943.
27. **Delmont TO, Robe P, Clark I, Simonet P, Vogel TM.** 2011. Metagenomic comparison of direct and indirect soil DNA extraction approaches. J Microbiol Methods **86:**397–400. http://dx.doi.org/10.1016/j.mimet.2011.06.013.
28. **Abbai NS, Govender A, Shaik R, Pillay B.** 2012. Pyrosequence analysis of unamplified and whole genome amplified DNA from hydrocarbon-contaminated groundwater. Mol Biotechnol **50:**39–48. http://dx.doi.org/10.1007/s12033-011-9412-8.
29. **Darling ACE, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394–1403.
30. **Randau L, Schröder I, Söll D.** 2008. Life without RNase P. Nature **453:**120–123. http://dx.doi.org/10.1038/nature06833.
31. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res **33:**W686–W689. http://dx.doi.org/10.1093/nar/gki366.
32. **Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV.** 2007. Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. Biol Direct **2:**33. http://dx.doi.org/10.1186/1745-6150-2-33.
33. **Jahn U, Gallenberger M, Paper W, Junglas B, Eisenreich W, Stetter KO, Rachel R, Huber H.** 2008. *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. J Bacteriol **190:**1743–1750. http://dx.doi.org/10.1128/JB.01731-07.
34. **Randau L.** 2012. RNA processing in the minimal organism *Nanoarchaeum equitans*. Genome Biol **13:**R63. http://dx.doi.org/10.1186/gb-2012-13-7-r63.
35. **Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL.** 2013. Microbial genomic taxonomy. BMC Genomics **14:**913. http://dx.doi.org/10.1186/1471-2164-14-913.
36. **Reno ML, Held NL, Christopher J, Burke PV, Whitaker RJ.** 2009. Biogeography of the *Sulfolobus islandicus* pan-genome. Proc Natl Acad Sci U S A **106:**8605–8610. http://dx.doi.org/10.1073/pnas.0808945106.
37. **Lillestøl RK, Redder P, Garrett RA, Brügger K.** 2006. A putative viral defense mechanism in archaeal cells. Archaea **2:**59–72. http://dx.doi.org/10.1155/2006/542818.
38. **Di Giulio M.** 2013. Is *Nanoarchaeum equitans* a paleokaryote? J Biol Res **19:**83–88.
39. **She Q, Shen B, Chen L.** 2004. Archaeal integrases and mechanisms of gene capture. Biochem Soc Trans **32:**222–226. http://dx.doi.org/10.1042/bst0320222.
40. **Bolduc B, Wirth JF, Mazurie A, Young MJ.** 30 June 2015. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. ISME J http://dx.doi.org/10.1038/ismej.2015.28.