Field deployable automated nucleic acid extraction system enables rapid on-site measurement of microbial diversity of Yellowstone National Park hot springs

Jason M. Wood  
Jet Propulsion Lab
Camilla Urbaniak  
Jet Propulsion Lab
Ceth W. Parker  
Jet Propulsion Lab
Nitin Kumar Singh  
Jet Propulsion Lab
Season Wong  
Al Biosciences, Inc.
Arunkumar Arumugam  
Al Biosciences, Inc.
Dana Skorupa  
Montana State University
Ashlyn Hemmah  
Montana State University
Phoebe Laaguiby  
University of Vermont
Fathi Karouia  
Ames Research Center
Brent Peyton  
Montana State University
Scott Tighe  
University of Vermont

KASTHURI VENKATESWARAN (kjvenkat@jpl.nasa.gov)  
Jet Propulsion Lab

Research Article
Field deployable automated nucleic acid extraction system enables rapid on-site measurement of microbial diversity of Yellowstone National Park hot springs

Jason M. Wood¹, Camilla Urbaniak¹, Ceth Parker¹, Nitin Kumar Singh¹, Season Wong², Arunkumar Arumugam², Dana J. Skorupa³,⁴, Ashlyn Hemmah³, Phoebe Laaguiby⁸, Fathi Karouia⁵,⁶, Brent M. Peyton³,⁴,⁷, Scott Tighe⁸*, and Kasthuri Venkateswaran¹*

¹ Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
² AI Biosciences, Inc., College Station, TX, USA
³ Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA
⁴ Department of Chemical and Biological Engineering, Montana State University, Bozeman, MT, USA
⁵ Space Bioscience Division, NASA Ames Research Center, Moffett Field, CA, USA
⁶ Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA
⁷ Thermal Biology Institute, Montana State University, Bozeman, MT, USA
⁸ Advanced Genomics Lab, University of Vermont Cancer Center, Burlington, VT, USA

* Correspondence:
Kasthuri Venkateswaran
Senior Research Scientist
Biotechnology and Planetary Protection Group; M/S 245-104
4800 Oak Grove Dr., Pasadena, CA 91109
Tel: (818) 393-1481; Fax: (818) 3934176
E-mail: kjvenkat@jpl.nasa.gov

Scott Tighe
Technical Director
Advanced Genomics Lab
Vermont Integrative Genomics Lab
University of Vermont Cancer Center
Health Science Research Facility room 303
149 Beaumont Ave
Burlington Vermont 05405
Tel: (802) 656-2482; Cell (802) 999-6666
E-mail: scott.tighe@uvm.edu
<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Email</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>Jason M. Wood</td>
<td><a href="mailto:woodinsitu@gmail.com">woodinsitu@gmail.com</a></td>
<td>0000-0002-0847-4695</td>
</tr>
<tr>
<td>40</td>
<td>Camilla Urbaniak</td>
<td><a href="mailto:camilla.urbanik@jpl.nasa.gov">camilla.urbanik@jpl.nasa.gov</a></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Ceth Parker</td>
<td><a href="mailto:ceth.w.parker@jpl.nasa.gov">ceth.w.parker@jpl.nasa.gov</a></td>
<td>0000-0002-4328-490X</td>
</tr>
<tr>
<td>42</td>
<td>Nitin Kumar Singh</td>
<td><a href="mailto:nitin.k.singh@jpl.nasa.gov">nitin.k.singh@jpl.nasa.gov</a></td>
<td>0000-0001-5344-1190</td>
</tr>
<tr>
<td>43</td>
<td>Season Wong</td>
<td><a href="mailto:season.wong@aibiosciences.com">season.wong@aibiosciences.com</a></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Arunkumar Arumugam</td>
<td><a href="mailto:arumugam.arunkumar@gmail.com">arumugam.arunkumar@gmail.com</a></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Dana J. Skorupa</td>
<td><a href="mailto:dana.skorupa@montana.edu">dana.skorupa@montana.edu</a></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Ashlyn Hemmah</td>
<td><a href="mailto:ashlyn.hemmah@gmail.com">ashlyn.hemmah@gmail.com</a></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Phoebe Laaguiby</td>
<td><a href="mailto:plaaguiby@gmail.com">plaaguiby@gmail.com</a></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Fathi Karouia</td>
<td><a href="mailto:fathi@bmsis.org">fathi@bmsis.org</a></td>
<td>0000-0002-1142-9684</td>
</tr>
<tr>
<td>49</td>
<td>Brent M. Peyton</td>
<td><a href="mailto:bpeyton@montana.edu">bpeyton@montana.edu</a></td>
<td>0000-0003-0033-0651</td>
</tr>
<tr>
<td>50</td>
<td>Scott Tighe</td>
<td><a href="mailto:scott.tighe@uvm.edu">scott.tighe@uvm.edu</a></td>
<td>0000-0002-3988-0741</td>
</tr>
<tr>
<td>51</td>
<td>Kasthuri Venkateswaran</td>
<td><a href="mailto:kjvenkat@jpl.nasa.gov">kjvenkat@jpl.nasa.gov</a></td>
<td>0000-0002-6742-0873</td>
</tr>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Abstract

Background: Microbial diversity estimation relies heavily on the extraction of nucleic acids from complex sample matrixes, regardless of the location of the laboratory used to recover the DNA, including those on Earth or in space. Nucleic acid sample preparation steps are labor intensive and require the effective lysis of cells and recovery of pure inhibitor free purification prior to downstream applications. An automated system is desirable for field deployment because of ease of use and autonomous rapid results. This extends to use in remote filed locations by NASA for the in situ measurements of life. The goal of this study was to setup a field-deployable lab in a remote setting with limited resources and demonstrate the ability to process and sequence nucleic acid samples on-site. Microbial mat and sediment samples were collected from several hot springs in Yellowstone National Park (YNP), targeting the Five Sister Springs (FSS), spring LNN010, and Octopus Spring (OS). Samples were processed on site and analyzed for microbial diversity using the field-deployable lab instrument.

Results:

Microbial diversity measured with both the Illumina (short reads) and Oxford Nanopore Technology (ONT; long reads) sequencing systems was highly similar. In general, for both the FSS and OS hot spring sites, bacteria (>90%) dominated over archaea (<10%). Metagenomic results were binned to existing databases such that the taxonomic ID represents the closest known organism with a sequenced genome. With this caveat the presence of, the most common archaeal community member in both the mat and sediment was related to Candidatus Caldiarchaeum subterraneum. Among the bacterial members, Roseiflexus sp. RS-1 was found to be highly abundant in the mat environment, as already reported. Notably, however, organisms related to Bacterium HR17 was abundant several FSS/OS mat samples. The detection of this novel bacterial community member warrants additional research.

Conclusion:

This is the first study to deploy an automated nucleic acid extraction system in YNP) to rapidly (~8 to 9 hrs) measure the environmental microbiome of hot springs. The demonstration of a field deployable system for nucleic acids extraction from complex materials (mats and sediments) highlights its utility not only in extreme environments on Earth but its potential benefit to NASA and deployment in space, once performance under microgravity is refined. The benefits of this new field-deployable lab include sample processing (extracting nucleic acids) with and without the use of the high performance cell wall degrading enzyme metapolyzyme, shotgun metagenome library preparation, and sequencing utilizing the space-tested Nanopore sequencing platform.

Keywords:
Automated, µTitan, environmental microbiome, microbial diversity, Yellowstone National Park
Introduction

Monitoring microbial burden is one of the environmental monitoring criteria needed by NASA to address identified risks to crew health during Exploration-class missions. Current utilization of biological testing in the International Space Station (ISS) is limited by available upmass (mass going up to space), downmass (mass coming back to earth from space), and crew time, as well as by the capabilities of the interfaces and hardware already developed. Other factors include reliance on return sample and ground analysis, constraints on size, mass, and power, lack of portability, and insufficient calibration life of current commercial systems. While a molecular detection system and Oxford Nanopore Technology (ONT) sequencer are available on board the ISS, the bottleneck remains labor intensive nucleic acid (NA) sample preparation, which mainly involves DNA/RNA isolation and purification steps.

Subsequent to a lab-based microbial diversity analysis using a whole-cell microbial reference community that yielded better recovery than commercial systems [1], the field-deployable μTitan system was tested at Yellowstone National Park (YNP). Commercially available nucleic acid extraction systems cannot be deployed to the field due to large footprints, large power requirements, and reagent stability to operate in zero gravity. The μTitan system is portable with minimal power and stability requirements, allowing operations at YNP extreme hot spring environments with limited field resources. The hot springs within areas of the lower geyser basin of YNP have been well studied [2-4] and offer multiple types of communities to sample that include both Bacteria and Archaea, including high-biomass microbial mats and sediments.

DNA/RNA based analysis in biodetection relies heavily on the success of the extraction of NA from complex sample matrixes, regardless of whether the detection is done in laboratories on Earth or in space [5-9]. Upstream sample preparation steps require the user to effectively lyse cells, recover NA, and purify NA by removing interfering contaminants. The samples may possess numerous contaminants (interfering or inhibiting components) depending on the sample type. While researchers can use manual sample preparation kits, these methods are labor intensive, susceptible to contamination, handling variations, or user error [10, 11]. An automated system in space is more desirable because of throughput and minimized crew-time involved and currently no system is designed for zero gravity.

Originally supported by the NASA Small Business Innovation Research program funding, with additional support from NASA’s Translational Research Institute for Space Health (TRISH) program, AI Biosciences, Inc. (AIBI) has developed a versatile, compact prototype called the μTitan and demonstrated its ability for rapid, medium-throughput automated sample processing. The goal for the deployment of the μTitan system in a remote site is to evaluate performance and real-time processing similar to what could be encountered in microbial monitoring on the ISS or Exploration-class missions.

An overview of the μTitan system has been thoroughly described previously [1]. The key feature of the μTitan is a self-contained enclosed NA extraction cartridge to perform DNA extraction in zero and standard gravity. The movement inside the closed cartridge is enabled by magnetic coupling with no direct contact with the sample. Magnetic particle-based DNA/RNA extraction is performed inside the cartridge but has the flexibility to use a range of extraction chemistries based on target molecule, or study requires aboard the ISS or other space transport vehicles.
The goal of this study was to evaluate field-deployable laboratory in a remote setting with limited resources and demonstrate the ability to process and sequence samples. Thus, µTitan combined with enzymatic pre-processing using DNA-free metapolyzyme and subsequent shotgun microbiome analysis using ONT field sequencing, and remote data analysis was performed while in the field.

**Results**

Microbial mat (n=15) and sediment (n= 33) samples were collected from numerous hot springs in Yellowstone National Park, including Octopus Spring, spring LNN010, multiple pools from Five Sisters Springs, and along their effluent channels (Figure 1). Detailed descriptions of the temperature and pH of the hot springs sampled are given in Table 1. In total, 48 samples were collected, including nine sampling device controls for all of the assays.

**Chemistry**

Vent fluid temperatures ranged between 68°C to 85°C, with the effluent channels of Octopus Spring and Five Sisters Springs measuring between 62°C to 64°C (Supplemental Table ST1). Aqueous chemistry was measured from water samples taken either directly from the vent fluid or above the photosynthetic mats. The White Creek thermal area is generally categorized as a water-dominated alkaline-chloride system [12], which is fed from subsurface boiling in deep-heated reservoirs. Collectively, the sampling results all feature high chloride (309 ± 10 mg/L), high arsenic (716 ± 29 µg/L) and low sulphate (18 ± 1 mg/L) concentrations (Supplemental Table ST1). In addition, the sampled springs are alkaline (pH 7.99–8.85) and contain high concentrations of sodium (140 ± 6 mg/L Na⁺), which is typical of the sodium and chloride-rich waters that form alkaline-chloride hot spring fluids.

**Microbial diversity**

Read counts for each detected species were combined for replicate samples collected from the same hot spring location before calculating relative abundances. In general, Nanopore data yielded >3kb long reads and Illumina yielded 150-bp reads.

**Archaeal diversity of mat samples**

Differential archaeal diversity of the microbial mats was assayed using the Illumina (Supplemental Table ST2) and ONT sequencing platforms (Supplemental Table ST3) with the primary findings presented in Figure 2 and Table 2. Approximately 67K of the Illumina short reads and ~2K of the ONT long reads were identified as Archaea after high quality trimming. In both the low-biomass and high-biomass microbial mat samples Candidatus Caldiarchaeum subterraneum was overwhelmingly present. It is interesting to note that a close relative of Haladaptatus pauchhalophilus (a halophilic archaeon) was detected at high relative abundance in FSSe1-H, FSSe1-L, and Ose, but not in FSSe2, by both the Illumina short read and ONT long read platforms. Whereas Pyrobaculum sp. WP30 (~12%) was found in high numbers at FSSe2-L mat samples. The community composition of the Octopus Spring mat resembled that of FSSe1 with respect to archaeal abundance and taxa profile.

The ONT long reads of Five Sisters Spring high biomass mat samples exhibited six archaeal taxa
from FSSe1 location (FSSe1-H) that were closer to the hot water run off (81.3°C, FSS3 and 78.5°C, FSS1), whereas a high number of Archaea (20 taxa) were observed in FSSe2 high biomass (FSSe2-H) samples that were collected from the cooler geothermal outflow channel (63.5°C; FSSe2). When the low biomass samples were analyzed, FSSe1 location (FSSe1-L) had 9 archaeal taxa whereas in Location 5 (FSSe2-L) samples only 2 archaea were retrieved. In FSSe1-H and FSSe2-H samples, the archaeal diversity was similar, with 64% to 71% of the archaeal diversity comprised of Candidatus C. subterraneum. It is also interesting to note that the archaeal taxa detected in the FSSe1 mats were all present at >1% abundance, while in the FSSe2-H mat, only 8 of the 20 archaeal taxa displayed a >1% abundance. The two archaeal taxa retrieved from the FSSe2-L mat samples belonged to Candidatus C. subterraneum (~46%) or Pyrobaculum sp. WP30 (~54%). The archaeal diversity profile of the Octopus Spring mat samples resembled the FSSe1-H and FSSe2-H archaeal diversity which included Candidatus C. subterraneum (~68%), archaeon HR02 (~12%), and HR01 (~11%) sequences constituting >91% of total archaeal abundance.

Bacterial diversity of mat samples

Illumina short reads of Five Sisters Spring high biomass mat samples exhibited 55 bacterial taxa from FSSe1 location (65.2 °C; FSSe1-H) that is closer to the hot water run off (81.3°C), whereas high number of bacteria (115 taxa) were observed in FSSe2 high biomass (FSSe2-H) samples that were collected from the relatively cooler area (63.5°C; FSSe2). Differential bacterial diversity of the microbial mats was assayed using the Illumina (Supplemental Table ST4) and ONT sequencing platforms (Supplemental Table ST5) with the primary findings presented in Figure 2 and Table 3. Illumina short reads of Five Sisters Spring high biomass mat samples exhibited 55 bacterial taxa from FSSe1 location (FSSe1-H) that were closer to the hot water run off (81.3°C, L1 and 78.5°C, FSS1), whereas high number of bacteria (115 taxa) were observed in FSSe2 high biomass (FSSe2-H) samples that were collected from the relatively warm area (63.5°C; FSSe2) (Figure 2; Table 3; Supplemental Table ST4). When low biomass samples were analyzed, FSSe1 location (FSSe1-L) had 76 bacterial taxa and in Location 5 (FSSe2-L) mat it was 103. Even though the bacterial diversity of FSSe1 resembles as that of FSSe2, their proportions were different as documented in ONT sequencing platform (Table 3). For example, Roseiflexus sp. RS-1 was the most abundant bacterial taxa detected in each of the mat samples; however, it’s overall abundance varied within each mat samples, ranging from 11-47%. Even though the bacterial population of FSSe2 mat was highly diverse (115 in FSSe2-H and 105 in FSSe2-L), all the highly prevalent taxa (present at >1% abundance) were similar to those detected in the other five sampled mats. Illumina short reads also retrieved lesser abundance of Thermocrinis ruber in FSSe1 (0.5%) mat when compared to FSSe2 (~3 to 4%) mat as that of ONT sequencing platform. Two additional bacterial taxa that were abundant from both FSSe1 mat samples belonged to Synechococcus sp. JA-2-3B’a(2-13) (14% in FSSe1-H and 13% in FSSe1-L) and Candidatus Thermochlorobacter aerophilum (7% in FSSe1-H and 5% in FSSe1-L samples). Similar incidences of these bacterial taxa at about 10% or more was noticed in FSSe2 mat samples as well.

The bacterial diversity profile of the Octopus Spring mat resembled FSSe1 and FSSe2 which included Roseiflexus sp. RS-1 (~34%), Synechococcus sp. JA-2-3B’a(2-13) (~23%), and Candidatus T. aerophilum (~8%) sequences constituting ~64% of total bacterial abundance. Even though 84
bacterial taxa were found from Octopus Spring mat samples, only 15 taxa exhibited >1% of the total bacterial reads.

The ONT long reads of Five Sisters Spring high biomass mat samples exhibited 89 bacterial taxa from FSSe1 location (FSSe1-H) that were closer to the hot water run off (81.3°C, FSS3 and 78.5°C, FSS1), whereas high number of bacteria (187 taxa) were observed in FSSe2 high biomass (FSSe2-H) samples that were collected from the relatively warm area (63.5°C; FSSe2) (Table 4; Supplemental Table ST5). When low biomass samples were analyzed, FSSe1 location (FSSe1-L) had 74 bacterial taxa whereas in FSSe2-L samples only 38 bacterial taxa were retrieved. Even though the bacterial diversity of FSSe1 mat samples resembles that of FSSe2 mat samples their proportions were different. Similar to the Illumina dataset, *Roseiflexus* sp. RS-1 was most abundant in all four mat samples analyzed via ONT, but the abundance levels of this community member varied, ranging between 5-40%. As well, lower abundance levels of *T. ruber* was detected in FSSe1-H (1%) mat samples when compared to FSSe1-L (8%), FSSe2-H (5%), and FSSe2-L (~34%). The two bacterial taxa retrieved from both FSSe1 mat samples belonged to *Synechococcus* sp. JA-2-3B’a(2-13) (22% in FSSe1-H and 13% in FSSe1-L) or *Candidatus* T. aerophilum (11% in FSSe1-H and 9% in FSSe1-L samples). Similar incidences of these bacterial taxa at about 10% or more was noticed in FSSe2 mat samples as well. The bacterial diversity profile of the Octopus Spring mat samples resembled the FSSe1-H and FSSe2-H bacterial diversity which included *Roseiflexus* sp. RS-1 (~30%), *Synechococcus* sp. JA-2-3B’a(2-13) (~29%), and *Candidatus* T. aerophilum (~11%) sequences constituting ~70% of total bacterial abundance. Even though 107 bacterial taxa were found from Octopus Spring mat samples, only 12 taxa exhibited >1% of the total bacterial reads.

Archaeal diversity of sediment samples

Differential archaeal diversity of sediment samples assayed using Illumina (Supplemental Table ST6) and ONT sequencing (Supplemental Table ST7) are presented (Figure 3 and Table 4). Approximately 1.2M Illumina short reads and ~55.9K ONT long reads associated with archaea after high quality trimming from the sediment samples were retrieved. Illumina short reads of sediment samples from FSS5-H, FSS1-H, and FSS2-H had 64, 15, and 31 archaeal taxa, respectively. Similarly, low biomass sediments from FSS5-L, FSS1-L, and FSS2-L possessed 106, 33, and 16 archaeal taxa, respectively. As shown in ONT long reads, Illumina also showed that high abundance of archaeal taxa was present in relatively higher temperature sediments collected from FSS5 (77.7°C) when compared to FSS1 (65.2°C) and FSS2 (63.5°C) sediment samples. The archaeal diversity observed in Illumina data was almost similar to that of ONT platform with abundant *Candidatus* C. subterraneum (30% to 39%), but *H. paucihalophilus*, a halophilic archaeon was absent in high temperature FSS5 sediment samples (77.7°C) when compared to FSS1 and FSS2 sediments of both high and low biomass sediment samples. The Octopus Spring sediment samples also resembled that of Five Sister Spring sediments with respect to archaeal abundance and taxa profile.

ONT long reads of Five Sisters Spring high biomass mat samples exhibited 65 archaeal taxa from FSS5 location (FSS5-H) that were closer to the hot water run off (81.3°C, FSS3 and 78.5°C, FSS1), whereas only 9 and 23 taxa were observed in FSS1 and FSS2 high biomass sediment samples that were collected from the relatively warm area (65 to 63°C). In high biomass sediment (FSS5,
FSSe1, and FSSe2) samples, the archaeal diversity was similar and high abundance of *Candidatus* C. subterraneum (42% to 70%) was noted. The second and third dominant taxa were archaeon HR02 and HR01 with ~10% of total archaeal abundance. As noticed in mat samples *Pyrobaculum* sp. WP30 was absent in sediment samples but this archaeon was dominant in the upstream water (~55% to 64%). The archaeal diversity profile of the Octopus Spring sediment samples resembled some Five Sister Spring results from FSS5, FSSe1, and FSSe2archaeal diversity which included *Candidatus* C. subterraneum (~57%), archaeon HR02 (~15%), and HR01 (~11%) sequences; constituting >84% of total archaeal abundance. In contrast, the hottest sediments from the LNN010 possessed extremely different archaea (*Ignisphaera aggregans*, *Thermoprotei* archaeon, and *Desulfurococcaceae* archaeon AG1) but in total, only 51 reads of ONT long reads were retrieved.

**Bacterial diversity of sediment samples**

Differential bacterial diversity of sediment samples assayed using Illumina (Supplemental Table ST8) and ONT sequencing (Supplemental Table ST9) are presented (Figure 3 and Table 5). Illumina data of Five Sisters Spring high biomass sediment samples exhibited 54 bacterial taxa from FSS5 location (FSS5-H) that were closer to the hot water run off (81.3˚C, FSS3 and 78.5˚C, FSS1), whereas high number of bacteria (115 taxa) were observed in FSSe2 high biomass (FSSe2-H) samples that were collected from the relatively warm area (63.5˚C; FSSe2) (Table 5). The sediment samples collected in between FSS5 and FSSe2 revealed 71 bacterial taxa (FSSe1-H; 65.2˚C). When low biomass samples were analyzed, FSS5 had 101 bacterial taxa, 118 taxa in FSSe1, and it was 87 taxa in FSSe2sediments; but their bacterial diversity was different. Bacterium HR17 (56%; 1.4M reads) was dominant in FSS5-L, *Synechococcus* sp. JA-2-3B’a(2-13) (15%; 0.23M reads) in FSSe1-L, and *Roseiflexus* sp. RS-1 (46%; 0.58M reads) in FSSe2-L sediments.

The bacterial diversity profile of the Octopus Spring sediment samples indicated the presence of 120 taxa among the total 2.87M reads retrieved and among them, only 22 taxa exhibited >1% of the total bacterial reads. The predominant bacterial taxa (>120K reads) of the Octopus Spring sediment included *Synechococcus* sp. JA-2-3B’a(2-13) (27%), Armatimonadetes bacterium CP1_7O (6%) and *Thermus aquaticus* (4%). The LNN010 sediment had only three recognizable bacteria constituting 3 bacterial taxa (58.4K reads) in which *T. ruber* (79.3%) was dominant.

ONT long reads of Five Sisters Spring high biomass sediment samples exhibited 96 bacterial taxa from FSS5 location (FSS5-H) that were closer to the hot water run off (81.3˚C, FSS3 and 78.5˚C, FSS1), whereas FSSe1-H and FSSe2-H sediment samples had only 6 and 23 bacterial taxa whose locations were relatively in warm area (66.5˚C and 63.5˚C). In contrast, when low biomass samples were analyzed, the bacterial taxa incidence was reversed where FSSe2-L (286 taxa) and FSSe1-L (142 taxa) had more compared to FSS5-L (85 taxa) sediment samples. The dominant bacteria present in these sediment samples were distinct in which FSS5-H and FSS5-L possessed bacterium HR17 (>66%) and HR19 (>12%) in high numbers but FSSe1 (3% to 7%) and FSSe2 (<1%) sediment samples had much lower abundance. On the contrary, FSSe1 sediments harbored high number of *Synechococcus* sp. JA-2-3B’a(2-13) (12% to 23%) and in FSS5 sediments *Roseiflexus* sp. RS-1 was abundant (31% to 32%). The sediment sample bacterial
diversity was not similar to the mat diversity. Unlike mat samples, the bacterial diversity profile
of the Octopus Spring sediment samples was different and had *Synechococcus* sp. JA-2-3B’a(2-
13) (27%) in abundance followed by bacterium HR36 (8.5%), *T. ruber* (7.8%), and bacterium
HR17 (5.1%). Even though 155 bacterial taxa were found from Octopus Spring sediments
(Supplemental Table ST9), only 20 taxa exhibited >1% of the total bacterial reads. The LNN010
sediments showed only 43 ONT long reads, of which only the recognizable member was *T.
ruber* and others were mostly attributed to “kitome” contaminants.

The three negative sampling device control samples were evaluated for exogenous DNA often
referred to as “kitome” and resulted in from 24.5K Illumina reads consisting of
*Stenotrophomonas* (55%), *Clostridioides* (19%), *Lacticaseibacillus* (12%), *Ralstonia* (12%), and
*Clostridium* sp. C105KSO13 (2%). The ONT data revealed only one bacterial taxon, *Roseiflexus*
sp, which maybe incidental contamination from the air by the sampling device (spatula) since
this bacterial taxon was found in abundant quantity in most of the hot spring samples.

*Illumina vs ONT methods*

Relative abundance of Illumina and Oxford Nanopore sequence data was similar among
samples collected from the same location with some differences. In general, both sequencing
platforms retrieved sequences associated with dominant bacterial and archaeal taxa (>95% of
the reads).

Analysis of similarities (ANOSIM) among similar sample types sequenced by either Illumina or
ONT demonstrate a high degree of similarity in relative sequence abundance. Mat samples
collected from the effluent channels of Octopus Spring and Five Sisters Springs resulted in a
similar set of microbial constituents sequenced (R=0.1652; P=0.015) by these two platforms
(Figure 4). Likewise, sediment samples collected from various sites, including Five Sisters
Springs and their effluent channel, the effluent channel of Octopus Spring, and LNN010
contained a similar set of microbes (R=0.1258; P=0.007). However, Illumina sequence data
indicated higher alpha diversity using the Chao1, Shannon, and Simpson indices for all sample
types (Figure 5). In addition, NMDS analyses also supported the differential microbial diversity
profiles of the examined hot spring samples where mat, sediment, and control samples
grouped themselves and exhibited differential profiles between the groups (Figure 5).

*Field measurement of microbial diversity using Nanopore*

In order to demonstrate the portability of the µTitan platform and the ONT sequencer, a field
measurement of microbial diversity of a single microbial mat sample collected from the effluent
channel of Five Sisters Spring (FSSe2) was performed while on site. During transport back to
Bozeman, MT, sequence data were generated using the MinION sequencer linked to a
computer and cellular. Data analysis was performed using the One Codex microbiome analysis
software. These immediate data (Supplemental Table ST10) found *Roseiflexus, Synechococcus,*
*Candidatus Thermochlorobacter*, and *Thermomicrobium*, all expected of this same sample.

*Mobile research sequencing laboratory*

These results demonstrate that it is possible to setup a mobile research sequencing laboratory
(MRSL) in a remote setting to rapidly (~8-9 hours for sample collection, processing, sequencing,
and bioinformatic characterization) attain measurements of microbial diversity. This is the first study of the field deployable µTitan system for extracting nucleic acids from complex materials, including hot water, microbial mats, and sediments in a remote setting like Yellowstone National Park. Additionally, this is the first reported use of Metapolyzyme, a specialized hydrolase cell wall degrading enzyme mixture, to better lyse microbial organism for a more complete microbiome analysis. Importantly, we demonstrated that the µTitan system can be used with either Illumina or ONT sequencing platforms, resulting in very similar perspectives of the microbial community. The portability of the µTitan and ONT platforms open more avenues for studying microbial diversity in remote locations.

**Discussion**

Demonstrating feasibilities of field deployment of miniaturized, low-powered, and portable instruments are of interest to NASA since such devices are of use for future space missions. The portable µTitan system has minimal power and stability requirements, allowing for the measurement of microbial diversity of YNP extreme hot spring environments with limited field resource. By coupling the µTitan with downstream shotgun metagenome analysis, we were able to measure an in-depth understanding of microbial dynamics in these extreme hot spring systems at species or strain level [13].

This is the first study using an automated nucleic acid extraction system deployed in a remote field such as YNP to rapidly measure an environmental microbiome. This workflow included various methods associated with sample collection, processing, extracting nucleic acids with, for the first time, the inclusion of the enzyme cocktail, Metapolyzyme. DNA sequencing operations were also included in the field using the MinION ONT platform, and remote data analysis performed via a cellphone hotspot. To our knowledge, this is the first study to compare shotgun microbiome data collected using Metapolyzyme assisted automated DNA extraction using both Illumina and ONT sequencing to benchmark the microbiome data results of each platform in measuring the microbial diversity of the YNP hot spring systems.

Previous studies have revealed that when low-biomass was extracted using µTitan system, the negative control “kitomes” of the reagents were distinct from the microbial diversity of samples [1] and replicate extractions (n=3) were consistent and reproducible and little difference between field or lab performance. In this study, operations were performed in the open field, surrounded by hot springs and steam clouds in the air during sample collection which can cause ubiquitous contamination of sampling gear with shorter fragment eDNA which may not be detected by long reads sequencers such as the ONT platform. This might be due to the fact that the microbiome of the air of these samples might have short sequences (picked by Illumina and not by ONT) that might have fragments of the hot spring microbes. This area of research needs more study.

The ability of Illumina and ONT to sequence small amplicons originating from clinical samples was performed and showed a genus-level concordance of 96.7% and 90.3% on the Illumina and ONT platforms, respectively [14]. Similarly, these two platforms were used to measure microbial diversity associated with bivalve communities, in Italy and Portugal and found that agreement between sequencing methods was only 69% (Illumina was higher) and the
difference between methods was non-significant [15]. In contrast, when indoor samples were assayed for 16S rRNA amplicons using Illumina and ONT platforms, there was no significant difference between the microbial compositions at the family level. However, at the genus, and particularly at the species level, the ONT MinION reported greater taxonomic resolution than Illumina MiSeq [16]. In this YNP study, as shown in Figure 5, no major differences were observed in the microbial diversity of mat and sediment samples when these two systems were used.

The waters of the FSS and OS hot springs are alkaline (~7.9 to 8.9 pH) in nature and high in chloride (300 to 340 mg/L), fluoride (27 to 30 mg/L), sulfate (14 to 18 mg/L), and arsenic (676 to 748 µg/L). In such an extreme environment, dominant microbes of certain microbial populations were documented based on targeted gene amplification/sequencing [17-19] and shotgun metagenome sequencing methods [20]. Though previous metagenomic studies have been conducted at Octopus hot spring [3], to date only 16S rRNA gene targeted amplicon sequencing has been used to discern the microbial community composition in the FSS hot spring system [21]. To our knowledge, work here is the first to generate shotgun metagenomics sequencing from the FSS system hot spring complex.

While detection of Candidatus C. subterraneum, an uncultivated Aigarchaeota has been reported in other microbial mats from geothermal sites including water stream of a sub-surface gold mine in Japan [22], Lobios hot spring water (76°C, pH = 8.2), Spain [23], Mid-Cayman Rise Vent Fluids [24], and in a Jinze hot spring in China (GenBank accession #: DTCM00000000.1), it was not previously been reported in the hot springs studied during this study until now. This may be attributed to the use of Metapolyzyme combined with deep shotgun sequencing. Since it is well known that this enzyme cocktail has remarkable cell wall degradation characteristics and that shotgun sequencing is far more agnostic and sensitive than 16S, it can potentially allow for more comprehensive lysis and detection, revealing entirely new microbial diversity. The metabolic and physiological features mined from the metagenome assembled genome of this uncultivated Aigarchaeota appeared to support aerobic conditions for growth with a heterotrophic lifestyle. Because of the reference database used here, metagenomic sequences detected from the microbial mats from the YNP system binned to Candidatus C. subterraneum in the microbial mats from the YNP system, however are likely more related to a different novel Aigarchaeota, Candidatus Calditenuis aerorheumensis previously isolated from OS [25]. In sediments, Candidatus C. subterraneum was overwhelmingly present as noted in mat samples. The LNN010 sediments contained 56 archaeal taxa as observed by short read Illumina sequencing but only 3 archaea were retrieved by the ONT platform. This discrepancy could be attributed to the fact that these archaeal taxa might not be viable and the genetic sequences retrieved by the Illumina system might be fragmented hence the long read ONT platform did not retrieve them.

It is interesting that Haladaptatus paucihalophilus was detected in the effluent channels of FSS and OS as these systems do not have the necessary salt concentration (0.8-5.1M; Supplemental Table ST1), pH (5-7.5; Table 1), or temperature (25-45 °C; Table 1) to grow this archaeon [26]. While the detection of this microorganism may simply be an artifact of inadequate reference
sequences in the database used for annotation, there is the possibility that there is a closely
related non-halophilic, thermophilic archaeon that has heretofore been undetected in these
hot spring systems. Halophiles are known to be difficult organisms to lyse using the standard
bead-beating method [27], and as such have likely been overlooked in prior studies.
Metapolyzyme used in this study is able to create spheroplasts quite easily, allowing for the
detection of this potentially novel archaeon. More studies are warranted.

Among bacterial members, sequences similar to Roseiflexus sp. RS-1 were found to be high in
FSS/OS mat samples during this study which was also reported in OS mats [28]. The members of
the Chloroflexota phylum, thermophiles capable of chlorophototrophy, were the major
components of photosynthetic microbial mats in sulfidic and non-sulfidic hot springs in
Yellowstone National Park, Wyo. (Nubel et al., 2003). It was reported that Roseiflexus
dominated in non-sulfidic hot springs and Chloroflexus sequences were retrieved from sulfidic
springs on the basis of genetic and lipid biomarker evidence [29, 30]. However, sulfate content
was high (14 to 18 mg/L) in FSS/OS systems where Roseiflexus was dominated when compared
to Chloroflexus among archaeon found which beg the question whether FSS/OS systems were
non-sulfidic or not.

Attempts to isolate Aigarchaeota and Candidatus C. subterraneum have thus far been
unsuccessful [25, 31]. Abundance of sequences related the Candidatus C. subterraneum in
sediment samples was similar to mat samples indicating that this archaeon might accumulate
and thrive in mat or sediments where they may depend on other microbial communities for
growth. However, differential bacterial abundance in sediments (dominated by Bacterium HR17
not Roseiflexus) in some FSS/OS mat samples need more research. Cultivation of Bacterium
HR17 in the laboratory and/or metagenome assembled genome of this uncultivated bacterium
would help to understand the significance of Bacterium HR17 in sediment samples of FSS/OS
systems.

**Contribution to the field**

This is the first study where automated nucleic acid extraction system was deployed in a
remote field like YNP and rapidly (~8 to 9 hrs) measured environmental microbiome.
Demonstration of field deployable system to extract nucleic acids from complex materials (hot
water, mats, and sediments) showed its usage not only for Earth extreme environments but
could be beneficial to be employed in space by NASA once issues with microgravity are
perfected in the future. Some of these benefits include sample processing (extracting nucleic
acids), shotgun metagenome library preparation, and sequencing utilizing microgravity-tested
ONT platform. This is the first reported use of Metapolyzyme, the specialize hydrolase cell wall
degrading enzyme mixture, to increase microbial lysis for microbiome analysis.

**Materials and Methods**

To ensure the safety of on-site personnel, sampling locations away from known thin-crust zones
were initially identified by experienced personnel, and sample collection was performed at a
safe distance from the edge with extendable poles.
All field experiments were conducted in YNP at alkaline geothermal features in the White Creek thermal area located in Lower Geyser Basin on June 18, 2019. Spring water, microbial mat, and sediment were collected from the Five Sisters Springs (44.5325 °N, 110.7971 °W), spring LNN010 (44.5325 °N, 110.7974 °W), Octopus Spring (44.5340 °N, 110.7978 °W), and effluent channels from these features (see Table 1 for a list of samples). The temperature and pH of each site was measured *in situ* using a HQ30d combined pH-temperature probe (HACH, Loveland, CO, USA), with dissolved concentrations of oxygen also measured in the field using either a high or low-range dissolved oxygen method and a portable colorimeter (Hach DR900, Hach Co., Loveland, CO).

**Aqueous Chemical Analyses**

The White Creek thermal area was selected for this study because of the variety of source pools with chemosynthetic zones above the upper temperature limit for photosynthesis as well as several robust outflow channels with prominent photosynthetic zones containing microbial mats. To document environmental conditions, measurements of water temperature and aqueous chemistry were either measured *in situ* with portable equipment or preserved on-site and analyzed back at the lab. Aqueous geochemistry was assessed for each hot spring by filter sterilizing site water (0.22 µm) directly into sterile 15 mL conical tubes. Some filtered site water was acidified in the field with 5% trace metal grade nitric acid prior to transport and used for total dissolved metals analysis. Concentrations of total metals were measured using an Agilent 7500ce ICP-MS by comparing to certified standards (Environmental Calibration Standard 5183-4688, Agilent Technologies, Santa Clara, CA, USA) at the Environmental and Biofilm Mass Spectrometry Facility at Montana State University - Bozeman. Ion chromatography was used to determine concentrations of dominant anions. For this, non-acidified filtered samples were analyzed using a Dionex ICS-1100 chromatography System (Dionex Corp., Sunnyvale, CA, USA) equipped with a 25 µL injection loop and an AS22-4x250 mm anion exchange column, using an eluent concentration of 4.5 mmol·L⁻¹ sodium carbonate and 1.4 mmol·L⁻¹ sodium bicarbonate flowing at a rate of 1.2 mL·min⁻¹. Filtered samples collected in closed headspace scintillation vials were used for total carbon, total nitrogen, dissolved inorganic carbon, and non-purgeable organic carbon measurements with a Shimadzu total organic carbon analyzer (TOC-VCSH, Shimadzu, Japan) coupled to a chemiluminescence detector (TNM-1 total nitrogen unit) in the Environmental Analytical Laboratory at Montana State University. An overview of the aqueous geochemistry is given in Supplemental Table ST1; the following dissolved constituents were not detected: Be, Mg, Al, V, Cr, Mn, Fe, Co, Ni, Cu, Se, Ag, Cd, Ba, U, and Bi.

**Sample collection**

Microbial mat and sediment samples were aseptically collected using a small autoclaved metal scoop. Collected samples were placed into a 50 mL Falcon tube (Thermo Fisher Scientific, Waltham, MA, USA) until further processing. Microbial mat and sediment samples collected from Five Sisters Springs, spring LNN010, and Octopus Spring were taken from sites determined to be safe within reach of the shore of the spring. Samples collected from effluent channels of these springs were taken from the approximate center of the flow path of water. The small metal scoop used to collect these samples was rinsed with spring water between samplings.
Triplicate mat samples were collected from several high biomass and low biomass hot spring locations, and primarily focused on the Five Sisters Spring (FSSe1 and FSSe2) and Octopus Spring (OSe). In total, 15 mat samples were analyzed and in addition, three kitome controls were included. Sediment samples in triplicate from high biomass (n=9) and low biomass (n=9) of three locations from the Five Sisters Spring (FSS5, FSSe1, and FSSe2) and one location from the Octopus Spring (OSe; n=3) were collected. Sediment samples collected from three Five Sisters Spring locations (FSS5, FSSe1, and FSSe2) were divided into high biomass (H) and low biomass (L) categories based on the visual inspection. The temperature of sediment samples of FSS5 was hot (77.7˚C; but above 72 ˚C, approximate upper limit for photosynthesis at this pH) [32], though not much warmer than FSSe1 (65.2˚C) and FSSe2 (63.5˚C) locations. Sediments from Five Sisters Spring FSSe2 was designated as FSS5-H and FSS5-L; from the same site as where FSSe1 mat sample collected was named FSSe1-H and FSSe1-L; and from the same site as where FSSe1 mat sample and FSSe2 water sample collected was called as FSSe2-H and FSSe2-L (Figure 1). In addition, sediment samples from Octopus Spring (OSe; 66.5˚C; n=3) and the hottest area spring sampled, LNN010 (85˚C), were also collected in triplicate (LNN010; n=3). Furthermore, three field controls included in the mat samples were considered here as control and in total there were 24 sediment samples analyzed (Table 1).

Mat samples collected from two of the Five Sisters Spring locations (FSSe1 and FSSe2) were divided into high biomass (H) and low biomass (L) categories based on visual inspection to insure that high biomass did not overwhelm the chemistry used by the µTitan. As shown in Figure 1, FSSe1 was just after the Five Sisters Spring 5 water run-off whereas FSSe2 was situated closer to the creek. Importantly, the temperature profiles of FSSe1 (65.2˚C) and FSSe2 (63.5˚C) locations were similar. Of note, the µTitan was able to measure microbial diversity in both high and low biomass samples, with only minor differences in composition detected.

Sample processing

Samples were preprocessed by adding 1.2 mL of sterile PBS and microbial mat and sediment samples were individually placed into 2 mL Matrix Lysing E bead beading tubes (MP Biomedicals, Irvine, CA, USA) and bead beaten for 15 s using a battery-powered oscillating power tool (JobPLUS ONE 18 V multi tool with P246 console & P570 Head attachment, Ryobi, Anderson, SC, USA). Each tube was placed and fixed tightly on the flat end of the blade using blue painter’s tape. The device was run on full power for one minute at 20,000 oscillations per minute. 25 µl of 10 µg µl⁻¹ MetaPolyzyme (MAC4L, Millipore Sigma, St. Louis, MO, USA) was added to each tube and incubated for 1 hr at 35 °C, followed by another round of bead beating for 15 s. Tubes were then left to settle for 15 min, before aliquoting 100 µl of the supernatant into the µTitan system for DNA extraction.

µTitan DNA extraction

A commercially available magnetic particle–based nucleic acid isolation kit (NucliSENS Magnetic Particle Extraction Kit, bioMerieux, Durham, NC, USA) was purchased, and the reagents from this kit were used to pre-fill the µTitan cartridges. These reagents included Lysis buffer, magnetic particle solution, wash buffer #1, wash buffer #2, wash buffer #3, and elution buffer. The µTitan extraction protocol (e.g., volume, time of incubation, and number of repeated washing steps) is as follows: 100 µL of the pre-processed sample was added to the 2nd well of a
µTitan cartridge containing 400 µL of lysis buffer and incubated for 10 m. Eight microliters of NucliSENS magnetic particles, for capturing NAs, were then added to the lysed sample solution in the 2nd well. The magnetic particles were intermittently mixed for 10 m, allowing for NA to bind to the MPs. The cartridge was then secured on the µTitan system. An extraction tip mounted on the follower magnet set was placed on the 1st well of the cartridge. Wells # 3–7 contained wash buffer # 1 (600 µL and 250 µL), wash buffer # 2 (600 µL and 250 µL), and wash buffer # 3 (250 µL). Well # 8 contained Tris elution buffer (60 µL). Each pre-processed sample was run in triplicate. To ensure the quality of the work and to check for contamination, molecular-grade water was used for extraction during each run of the machine, instead of sample, and is considered the µTitan instrument control.

The µTitan system was operated in a vehicle in the parking lot using a laptop computer equipped with open-source software (Repetier-Hot-World GmbH & Co., Willich, Germany) that uses G-code based programing. Full operations of the µTitan system have been previously described [1].

**Sequencing**

Field sequencing was performed on a single sample of microbial mat collected from the effluent channel of Five Sisters Spring (FSSe2). DNA was immediately extracted in the field using the µTitan automated field sample extraction system[1] and quantified using a qubit spectrofluorometer using DNA high sensitivity reagents (Q32851, Thermo Fisher Scientific, Waltham, MA, USA). The concentration of recovered DNA was 0.5ng/µl and a total of 8 ul (4 ng) was used for the Rapid Sequencing Kit (SQK-RAD004, Oxford Nanopore Technologies, Oxford, United Kingdom). While this is much lower than the recommended 200-400ng input, the expectation was fewer sequences generated. Library synthesis was carried out in a parking area at Yellowstone National Park that we termed Mobile Research Sequence Laboratory (MRSL; Figure 1) and a portable heat block was used for the two incubation steps of 30°C and 80°C. Sequencing was conducted using a standalone version of the MinKNOW software specially arranged for NASA field sequencing with the ONT MinION MK1B [1]. After flow cell QC was complete, the RAD004 library was added to the R9.4 flow cell and sequencing was initiated. Sequencing was performed in the vehicle while driving back to Bozeman, Montana with intermediate data analysis performed via a cell phone hotspot connection and the One Codex online software. Sequencing was allowed to continue until no additional data was generated. A total of 4.6 mb were generated from 3843 reads with an average length of 1.2 kb.

Laboratory-based Oxford Nanopore sequencing was performed using the GridION-X5 sequencer equipped with Standard flow cells (R9.4). Libraries were synthesized using the rapid PCR Barcoding kit (RPB004) with 3 µl of adjusted input volume. PCR cycles were adjusted for total input amounts as follows; 16 cycles for 1-3 ng, 18 cycles for 0.25-1ng, and 22 cycles for 0.0 to 0.25ng. Read depth varied for each sample and low reads were obtained for samples that had DNA at or below detection limits, but not for all cases. Low read depth was noted for some samples that had high inputs and may be explained by possible microbial content, biofilm interference, or other possible inhibitors such as excessive extracellular polysaccharides or organic acids.
Whole genome shotgun library synthesis was performed on 1ng input or maximum input volume using the Nextera XT whole genome kit (Illumina Inc., San Diego, CA, USA), cleaned, quantified using the Qubit fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA), and analyzed for quality using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Libraries were pooled equimolar and sequenced using single-end 150 bp chemistry on a HiSeq 1500/2500 system (Illumina, San Diego, CA, USA). Samples were demultiplexed and fastq files were used for analysis.

Sequence Quality Control and Annotation

Quality control of short-read Illumina sequence data was performed using Trimmmomatic (v0.32) to remove adapter sequences and low-quality reads, using a quality cutoff value set at minimum Phred score of 20 along the length of the read. Quality control of long-read Nanopore sequence data was performed using Nanofilt (v2.6.0) and default parameters. Reads shorter than 80 bp were removed from the data. The DIAMOND aligner (v2.0.4) was used to query the NCBI-NR protein database, and the MEGAN 6 lowest common ancestor (LCA) algorithm was used to bin high quality reads to their respective taxonomy. Additionally, the eggnog, KEGG, and SEED databases were queried for functional analysis. For long-read Nanopore sequence annotation, the DIAMOND frameshift and range curling mode was used. The frameshift option performs frameshift-alignment of DNA sequences against a protein reference database. The parameter used with the frameshift mode was -F 15, leading to a dynamic programming penalty of 15. Range-culling estimates an alignment locally; this option was used with the parameter as follows --range-culling and --top 10.

A read threshold was set to understand the abundance of microbial diversity in these samples. For ONT long reads, a minimum of 10 reads was considered and for Illumina, >500 reads was considered for the microbial abundance calculations. When computed, >99.9% of total reads were accounted for with these thresholds.

Statistical Analyses

To determine whether long-read ONT sequencing could supplement or replace short-read Illumina sequencing for more rapid analyses in remote locations, multiple statistical analyses were performed. Analysis of Similarities (ANOSIM) using the R vegan package [33] was used to compare similarities in community composition among similar samples (e.g., mat samples, sediment samples) as sequenced using either the Illumina or Oxford Nanopore platforms. Alpha diversity among similar samples was measured with Chao1, Shannon, and Simpson indices as calculated by the R vegan package [33] and a custom R script (https://github.com/sandain/R/blob/main/diversity.R) that performed rarefication based on the size of the smallest library. The R vegan package [33] was also used to visualize beta diversity using nonmetric multidimensional scaling (NMDS) of Bray-Curtis dissimilarity.

Ethics approval and consent to participate

Not applicable.

Acknowledgements
This study was conducted under Yellowstone National Park research permit YELL-2018-SCI-5480 held by BMP. Adriana Gonzalez (student) and Bo Liu (A.I. Biosciences). We acknowledged the Jet Propulsion Laboratory supercomputing facility staff, notably Narendra J. Patel (Jimmy) and Edward Villanueva, for their continuous support in providing the best possible infrastructure for BIGDATA analysis. © 2023 California Institute of Technology, Government sponsorship acknowledged.

**Funding**

This research was supported by the TRISH through Cooperative Agreement NNX16AO69A awarded to KV. The W. M. Keck Foundation provided supported for DJS and BMP. The NASA postdoctoral fellowship supported part of JMW, CP, and CU time. Preliminary work of this research was supported by a NASA SBIR Contract (NNX17CP21P) awarded to SW. The funders had no role in the study design, data collection, and interpretation; the writing of the manuscript; or the decision to submit the work for publication.

**Availability of data and materials**

The raw sequencing data has been deposited in NCBI SRA under BioProject accession number PRJNA935810.

**Disclaimer**

This manuscript was prepared as an account of work sponsored by NASA, an agency of the US Government. The US Government, NASA, California Institute of Technology, Jet Propulsion Laboratory, and their employees make no warranty, expressed or implied, or assume any liability or responsibility for the accuracy, completeness, or usefulness of information, apparatus, product, or process disclosed in this manuscript, or represents that its use would not infringe upon privately held rights. The use of, and references to any commercial product, process, or service does not necessarily constitute or imply endorsement, recommendation, or favoring by the US Government, NASA, California Institute of Technology, or Jet Propulsion Laboratory. Views and opinions presented herein by the authors of this manuscript do not necessarily reflect those of the US Government, NASA, California Institute of Technology, or Jet Propulsion Laboratory, and shall not be used for advertisements or product endorsements.

**Authors’ contributions**

JMW, CU, CP, NKS, SW, AA, DJS, FK, BMP, ST, and KV were involved in planning the field expedition to Yellowstone National Park and performed various functions on-site, including sampling, sample processing, and nucleic acid extraction using the µTitan. JMW, DJS, and BMP provided local knowledge of the sampling site and its dangers. JMW and CP were trained to handle any emergency situations while on-site. DHS and AH contributed to the geochemistry components of this work. PL performed DNA library synthesis and DNA quality control. PL and ST performed DNA sequencing using the ONT and Illumina platforms. JMW and NKS performed bioinformatic analyses. BMP, ST, and KV supervised all work. JMW, CU, CP, NKS, SW, ST, and KV contributed to writing this manuscript. All authors have read the manuscript and agree with its content.

**Competing interests**
The authors declare that they have no competing interests.

Figure legends
Figure 1: Locations of the Yellowstone National Park hot spring samples collected.
Figure 2: Bacterial and archaeal diversity of YNP hot spring mat samples.
Figure 3: Bacterial and archaeal diversity of YNP hot spring sediment samples.
Figure 4: NMDS analysis of YNP hot spring mat and sediment samples.
Figure 5: Microbial diversity indices of YNP hot spring samples.

Table legends
Table 1. Yellowstone National Park hot spring samples collected during this study.
Table 2. Dominant archaeal sequences retrieved from YNP Springs mat samples.
Table 3. Dominant bacterial sequences retrieved from YNP Springs mat samples.
Table 4. Dominant archaeal sequences retrieved from YNP Springs sediment samples.
Table 5. Dominant bacterial sequences retrieved from YNP Springs sediment samples.
References


**Figure 1**

Locations of the Yellowstone National Park hot spring samples collected.
Figure 2

Bacterial and archaeal diversity of YNP hot spring mat samples.

Figure 3

Bacterial and archaeal diversity of YNP hot spring sediment samples.
Figure 4

NMDS analysis of YNP hot spring mat and sediment samples
Figure 5

Microbial diversity indices of YNP hot spring samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• TablesMar2023MatandSedimentonly.pdf
• SupplementalTablesMar2023MatandSedimentonly.xlsx