**Comparative Metagenomics at Solfatara and Pisciarelli Hydrothermal Systems in Italy Reveal that Ecological Differences Across Substrates are not Ubiquitous**

Ifeoma R. Ugwuanyi1\*, Marilyn L. Fogel2, Roxane Bowden3, Andrew Steele3, Giuseppe De Natale4,5, Claudia Troise4,5, Renato Somma4,6, Monica Piochi4, Angela Mormone4 and Mihaela Glamoclija1\*

1Department of Earth and Environmental Sciences, Rutgers University, Newark, NJ, United States.

2EDGE Institute, University of California, Riverside, CA, United States.

3Earth and Planets Laboratory, Carnegie Institution for Science, Washington DC, United States.

4Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano, Naples, Italy

5 Consiglio Nazionale delle Ricerche INO, Naples, Italy.

6 Consiglio Nazionale delle Ricerche IRISS, Naples, Italy.

**\*Correspondence:**

Mihaela Glamoclija

m.glamoclija@rutgers.edu

Ifeoma R. Ugwuanyi

ifeoma.ugwuanyi@rutgers.edu

**Materials and Methods**

**Sample Collection**

To investigate the near-surface microbial community composition, structure, and function at two hydrothermal systems, water, mud, and fumarolic deposits samples were collected from Solfatara and Pisciarelli (Fig. 1A and 1B), in October 2012. All samples were collected aseptically, using sterile Falcon tubes, Nalgene bottles, scoops, and gloves. The samples were collected from Solfatara, in areas that were not accessible to park visitors and from Pisciarelli, in areas that were not frequently traversed by locals. At Solfatara (N 40° 49' 38.35''; E 14° 8' 31.22''), partially encrusted fumarolic deposit samples (Fig. 1D) were collected about 15 m away from the main fumarole, Bocca Grande (Fig. 1C). At Pisciarelli, hot fumarolic deposit samples (Fig. 1H) were collected from the hill about 10 m away from the boiling hot mud pools (Fig. 1F) and several meters away from the Istituto Nazionale di Geofisica e Volcanologia (INGV) permanent gas CO2 soil degassing monitoring station. The areas where fumarolic deposits were sampled were dry, and they contained fumarolic emissions that formed small (few cm in diameter) holes in the ground. This vesicular texture, along with a mixture of hardened particles about 1 to a few centimeters in size, and a minor contribution of fine-grained material, created an overall very heterogeneous texture. Fumarolic deposits and corresponding environmental measurements were collected at about 1 cm below the surface. Due to the hardened heterogeneous texture of the deposits, removing the top centimeter may not have been a very efficient way to avoid environmental contaminants. The largest mud pool was selected as a second sampling point at Solfatara (Fig. 1E). Two distinct mud pools were sampled at Pisciarelli: a small light-colored (beige) pool (Fig. 1I; N 40° 49’ 44.75’’; E 14° 8’ 49.74’’) and a large dark gray pool (Fig. 1G; N 40° 49’ 45.65’’; E 14° 8’ 48.57’’). Additionally, the large pool had a mud outlet and discharge channel where we also collected mud samples. Mud samples from Solfatara, due to safety concerns were collected about 50 cm above the water level on the lower soft edge; this mud was not submerged in the water and contained vesicles from the previous periods of active gas emissions. Mud samples from Pisciarelli small mud pool were collected about 20 cm above the water level on the inside wall of the pool hole; this mud was not submerged in water but was soft and moist. Mud samples from the Pisciarelli large pool were collected on the edge of the water level; this sampling site was occasionally splashed with bubbling water and the mud was a bit waterier than mud from Solfatara or the small Pisciarelli mud pool. Mud from the outlet was collected on the edge of the channel, in the nature previously described for the mud sample from the large mud pool. The difference between mud samples from the large mud pool and outflow channel is the temperature; the samples were collected along a small transect to obtain an insight into local environmental gradients (water-mud-channel). Water samples were collected using a sterile 500 ml Nalgene bottle, and Falcon tubes were used to collect mud samples from the pool wall. Before collecting mud samples, best efforts were made to remove the top 1 cm of the mud layer. Temperature, pH, and redox potential (Eh) of bubbling mud and fumarolic deposits were measured *in situ* with a portable probe. Water measurements were collected from the first bottle that was sampled and the water was discarded. Mud measurements were collected about 1 cm below the surface. Green epilithic microbial layer samples were also collected from the dry mud wall about 1 m away from Pisciarelli large pool (Fig. 1J). This sample was scooped using a Falcon tube to include as much green epilith as possible. The measurements for the epilithic sample were difficult to obtain because on reaching the surface, hydrothermal gasses diffuse and mix vigorously with air, preventing the probe reading from stabilizing. As a result, measurements were collected close to the sampling point where gasses were moving less vigorously. All samples were collected in triplicates and stored at -20 °C until further processing; samples for long term storage were stored at -80 °C. Gas readings were taken from the continuous INGV gas monitoring station at Solfatara and Pisciarelli (Table 1).

**X-ray fluorescence (XRF) Analysis**

X-ray fluorescence (XRF) was used to examine the chemical composition of mud and fumarolic deposits. Before XRF analysis, samples were dried in an oven at 70 °C for 24 h and then crushed into a homogenous powder using an agate mortar and pestle. Samples were analyzed in triplicate using a Horiba XGT-1000WR X-ray Fluorescence with an Rh tube X-ray source at the Department of Chemistry, Rutgers University (Busch Campus). Operating conditions were under a partial vacuum using 50 kV, 100 uA, and 120 s scans. Elemental wt. % was determined using the XGT-1000WR software's quantification.

**DNA Extraction and Metagenome Sequencing**

About 0.25 g of mud and fumarolic deposits was used for DNA extraction. Before DNA extraction, fumarolic deposits were powdered using a sterilized agate mortar and pestle. Water samples were filtered in the laboratory using a 0.2 µm VWR black polycarbonate filter. Filters were cut into small pieces with a sterile scalpel and used for DNA extraction. DNA was extracted using DNeasy PowerSoil kit (Qiagen Inc, Valencia, CA) with modifications to the manufacturer’s instructions. The three modifications described below were tried on each sample but only one modification yielded DNA per sample. **Modification 1**)**Pre-bead beating:** sample was added to a clean bead tube containing bead solution, vortexed for 5 mins to detach microbial cells from solid particles, and then centrifuged at 10,000 rpm for 30 s. The supernatant was transferred to a clean bead tube containing bead solution and processed following the manufacturer’s instructions. **Modification 2) Sonication:**sample was resuspended in 1.5 ml of phosphate-buffered saline (PBS) and sonicated at room temperature for 25 s using an ultrasonic bath (Fisher Scientific, Pittsburgh, PA) at a frequency of 40 kHz to detach microbial cells from solid particles. The sample was centrifuged at 2000 rpm for 30 s to remove large particles. The resulting supernatant was transferred to a clean bead tube containing bead solution and processed following the manufacturer’s instructions [1]. **Modification 3**)**Heating:** sample was added to a clean bead tube containing bead solution and incubated at 95 °C for 15 mins. After incubation, the sample was vortexed briefly and centrifuged at 10,000 rpm for 30 s. This way, microbial cells are mechanically separated from mineral substrates and particles that inhibit DNA extraction. The supernatant was transferred to a clean bead tube and processed following the manufacturer’s instructions [2]. DNA extraction for each sample was carried out in triplicate to account for sample heterogeneity. Positive (topsoil) and negative (reagent-only) controls were used during each batch of DNA extraction to ensure extraction quality.

Due to low DNA yield, DNA extraction in samples and controls was confirmed using Polymerase Chain Reaction (PCR). The primer pair B27F (AGA GTT TGA TCM TGG CTC) and B1429R (GGT TAC CTT GTT ACG ACT T) [3, 4] was used to confirm the presence of bacteria. Primer pair A8F (TCC GGT TGA TCC TGC C) and A1492R (GGC TAC CTT GTT ACG ACT T) [5] was used to confirm the presence of archaea. The presence of eukaryote was confirmed using EUK-1A (5′ CTG GTT GAT CCT GCC AG 3′) and EUK-1776R (5′ CGG AAA CCT TGT TAC GAC 3′) [6, 7]. The thermal cycler conditions were 95 °C for 2 min, 32 cycles of 95 °C for 30 s, 49.4 °C for 30 s, 72 °C for 45 s, and a final 10 min extension at 72 °C. Negative control (nuclease-free water) was used during PCR to account for reagent and sample handling contamination. PCR products were visualized on the Agilent 2100 Bioanalyzer system (Agilent Technologies, CA) using the Agilent 7500 kit. Equal volumes of triplicate DNA were pooled, and one sample was randomly picked to be used as a sequencing replicate. The pooled DNA of the randomly chosen sample was split into two equal volumes and used as sequencing replicates. We had a total of 7 replicates that were sequenced: 6 sample replicates and 1 sequencing replicate. DNA samples were stored at -80 °C until further processing.

**Identification and Filtering of Contaminants**

Contamination skews microbial composition and leads to an overestimation of diversity, especially in low biomass samples [8]. Karstens, et al. [8] showed that the proportion of contaminants increased as the starting biomass decreased. Therefore, it is important to minimize, identify, and eliminate contaminants that may alter the community composition of low biomass hydrothermal samples like that of Solfatara and Pisciarelli. To avoid laboratory-derived contamination, DNA extraction and PCR preparation were carried out in an aseptic way. DNA extraction and PCR preparation were carried out in a precleaned laminar flow cabinet; cleaning involved wiping surfaces with 10% bleach and later 90% Isopropyl wipes, followed by UV irradiation for 15 mins. All pipette tips with filters and tubes were certified DNA-free. All forceps and filter manifolds were autoclaved before use. Negative (reagent-only) and positive (soil sample) controls were used during each batch of DNA extraction to ensure extraction quality. Each batch of DNA extraction contained samples from both Solfatara and Pisciarelli. Negative control from DNA extraction, PCR negative control (nuclease-free water), and positive control (DNA extracted from soil ) were included during PCR amplification steps. Primers were reconstituted using nuclease-free water. PCR products were visualized on the Agilent 2100 Bioanalyzer system (Agilent Technologies, CA) using the Agilent 7500 kit and only samples that had no visible band in DNA and PCR negative controls were sent out for sequencing.

Following taxonomic classification, potential contaminants were identified and removed from the taxonomy table before performing downstream analyses. Any genera previously identified as molecular biology kit and laboratory contaminants [9, 10], which hereinafter are referred to as “kitome”, were removed from our taxonomy table. All unclassified genera in our taxonomy table that belong to families listed as kitome were also removed. Additionally, we removed singletons and species present in only one sample, since these may be the result of sequencing errors that create false taxa [11-13]. Kitome, singleton, and species in only one sample (Table S4, S5) contributed between 0.2 – 50 % (Supplementary Fig. S2 and S3) of the estimated community with samples from Solfatara having the highest level of contamination possibly due to low amounts of starting material which may have been overwhelmed by trace amounts of DNA in reagents, different types of kits and various laboratory surfaces, including sequencing facilities [14, 15]. Sequences that were unclassified at the domain level were also removed from the taxonomy table. Species with < 0.01 % relative abundance in each sample were excluded from our dataset since they are likely not true members of the hydrothermal community but may be contaminants introduced from the surrounding environment or carryover from other samples during the DNA extraction process. After removing all the taxa that fall into the categories listed above, the resulting dataset was used for downstream analyses and relative abundance was recalculated to generate Fig. 3 and Fig. 4. Since it is not possible to eliminate contamination in low biomass samples, especially with the increasing sensitivity of sequencing technologies, the best approach for low biomass samples is to include quality control measures and remove potential contaminants before data interpretation, as was done in this study. A limitation of this study is the absence of sequences from negative controls. DNA extraction and PCR were carried out using aseptic techniques to avoid contamination, and we have also carefully excluded from our dataset any taxa that could potentially be a contaminant to avoid an overestimation of diversity. However, without sequences from the negative controls, it is hard to say with certainty that all contaminants were filtered from the dataset prior to performing downstream analysis.**References**

1. Sánchez-Andrea I, Rodríguez N, Amils R, Sanz JL (2011) Microbial diversity in anaerobic sediments at Rio Tinto, a naturally acidic environment with a high heavy metal content. Appl Environ Microbiol 77: 6085-6093.

2. Glamoclija M, Ramirez S, Sirisena K, Widanagamage I (2019) Subsurface Microbial Ecology at Sediment-Groundwater Interface in Sulfate-Rich Playa; White Sands National Monument, New Mexico. Front Microbiol 10: 2595.

3. DeLong EF (1992) Archaea in coastal marine environments. PNAS 89: 5685-5689.

4. Madden AS, Smith AC, Balkwill DL, Fagan LA, Phelps TJ (2007) Microbial uranium immobilization independent of nitrate reduction. Environ Microbiol 9: 2321-2330.

5. Teske A, Hinrichs K-U, Edgcomb V, de Vera Gomez A, Kysela D, Sylva SP, Sogin ML, Jannasch HW (2002) Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. Appl Environ Microbiol 68: 1994-2007.

6. Abbott CL, Gilmore SR, Lowe G, Meyer G, Bower S (2011) Sequence homogeneity of internal transcribed spacer rDNA in Mikrocytos mackini and detection of Mikrocytos sp. in a new location. Dis Aquat Org 93: 243-250.

7. Maza-Márquez P, González-Martínez A, Martinez-Toledo M, Fenice M, Lasserrot A, González-López J (2017) Biotreatment of industrial olive washing water by synergetic association of microalgal-bacterial consortia in a photobioreactor. Environ Sci Pollut Res 24: 527-538.

8. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S (2019) Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. MSystems 4: e00290-00219.

9. Sheik CS, Reese BK, Twing KI, Sylvan JB, Grim SL, Schrenk MO, Sogin ML, Colwell FS (2018) Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the census of deep life. Front Microbiol 9: 840.

10. Weyrich LS, Farrer AG, Eisenhofer R, Arriola LA, Young J, Selway CA, Handsley‐Davis M, Adler CJ, Breen J, Cooper A (2019) Laboratory contamination over time during low‐biomass sample analysis. Mol Ecol Resour 19: 982-996.

11. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS 108: 4516-4522.

12. Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytol 188: 291-301.

13. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10: 996-998.

14. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12: 1-12.

15. Stinson LF, Keelan JA, Payne MS (2019) Identification and removal of contaminating microbial DNA from PCR reagents: impact on low‐biomass microbiome analyses. Lett Appl Microbiol 68: 2-8.