



# Unraveling the microbial and functional diversity of Coamo thermal spring in Puerto Rico using metagenomic library generation and shotgun sequencing



Ricky Padilla-Del Valle, Luis R. Morales-Vale, Carlos Ríos-Velázquez \*

Biology Department, University of Puerto Rico at Mayaguez, 108 Street Bo. Miradero Km 1.3, Zoo Entrance, Mayagüez 00680, Puerto Rico

## ARTICLE INFO

### Article history:

Received 13 December 2016

Accepted 17 December 2016

Available online 23 December 2016

### Keywords:

Metagenome  
Shotgun sequencing  
Fosmid library  
Thermal spring  
Bioprospect  
Coamo  
Puerto Rico

## ABSTRACT

In Puerto Rico, the microbial diversity of the thermal spring (ThS) in Coamo has never been studied using metagenomics. The focus of our research was to generate a metagenomic library from the ThS of Coamo, Puerto Rico and explore the microbial and functional diversity. The metagenomic library from the ThS waters was generated using direct DNA isolation. High molecular weight (40 kbp) DNA was end-repaired, electro eluted and ligated into a fosmid vector (pCCFOS1); then transduced into *Escherichia coli* EPI300-T1<sub>R</sub> using T1 bacteriophages. The library consisted of approximately 6000 clones, 90% containing metagenomic DNA. Next-Generation-Sequencing technology (Illumina MiSeq) was used to process the ThS metagenome. After removing the cloning vector, 122,026 sequences with 33.10 Mbps size and 64% of G + C content were annotated and analyzed using the MG-RAST online server. Bacteria showed to be the most abundant domain (95.84%) followed by unidentified sequences (2.28%), viruses (1.67%), eukaryotes (0.15%), and archaea (0.01%). The most abundant phyla were *Proteobacteria* (95.03%), followed by unidentified (2.28%), unclassified from viruses (1.74%), *Firmicutes* (0.20%) and *Actinobacteria* (0.18%). The most abundant species were *Escherichia coli*, *Polaromonas naphthalenivorans*, *Albidiferax ferreducens* and *Acidovorax* sp. Subsystem functional analysis showed that 20% of genes belong to transposable elements, 10% to clustering-based subsystems, and 8% to the production of cofactors. Functional analysis using NOG annotation showed that 82.79% of proteins are poorly characterized indicating the possibility of novel microbial functions and with potential biomedical and biotechnological applications. Metagenomic data was deposited into the NCBI database under the accession number SAMN06131862.

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications	
Organism/cell line/tissue	Metagenomic library of thermal spring waters in Coamo, Puerto Rico
Sex	Not applicable
Sequencer or array type	Illumina MiSeq
Data format	Raw data: or FASTQ file
Experimental factors	Environmental sample
Experimental features	Metagenomic library and shotgun sequencing performed from water obtained from Coamo thermal spring in Coamo, Puerto Rico
Consent	Not applicable
Sample source location	Water sample, thermal spring, Coamo, Puerto Rico (18°02' 16.6"N 66°22'27.6"W)

\* Corresponding author at: University of Puerto Rico at Mayaguez, Biology Department, PO Box 9000 Mayaguez 00681-9000 Puerto Rico.

E-mail addresses: [ricky.padilla@upr.edu](mailto:ricky.padilla@upr.edu) (R. Padiilla-Del Valle), [luis.morales28@upr.edu](mailto:luis.morales28@upr.edu) (L.R. Morales-Vale), [carlos.rios5@upr.edu](mailto:carlos.rios5@upr.edu) (C. Ríos-Velázquez).

## Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/biosample/6131862>.

## 1. Introduction

Different oligotrophic water environments, such as oceans, rivers and thermal springs from hot vents have been the focus of metagenomics, specifically because of the microbial diversity contribution to the ecosystem's stability and its life maintaining. High temperature aquatic ecosystems such as thermal springs, harbor unique thermophilic and hyperthermophilic microorganisms [3,4,13]. Classic microbiology has studied environmental microorganisms using culture-dependent approaches. Emerging sciences have led to research evidencing that approximately 0.1%–1.0% of the microbial community can be cultivable. Metagenomics is a culture-independent approach based on the isolation of environmental genomic material to provide a comprehensive taxonomic and functional evaluation of all the collected microorganisms [8]. One of the advantages of performing metagenomic

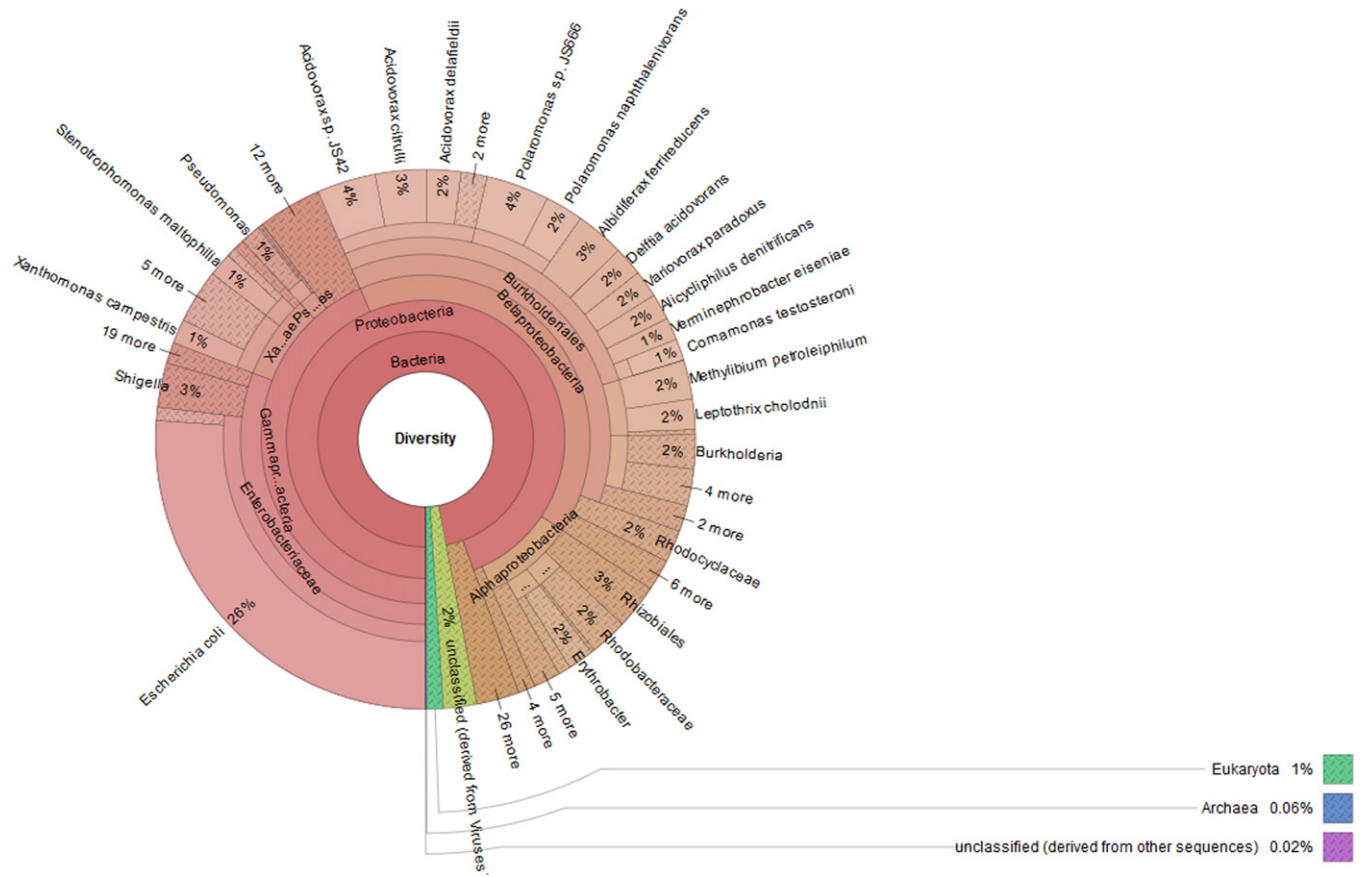


Fig. 1. Community structure of Coamo thermal spring metagenome.

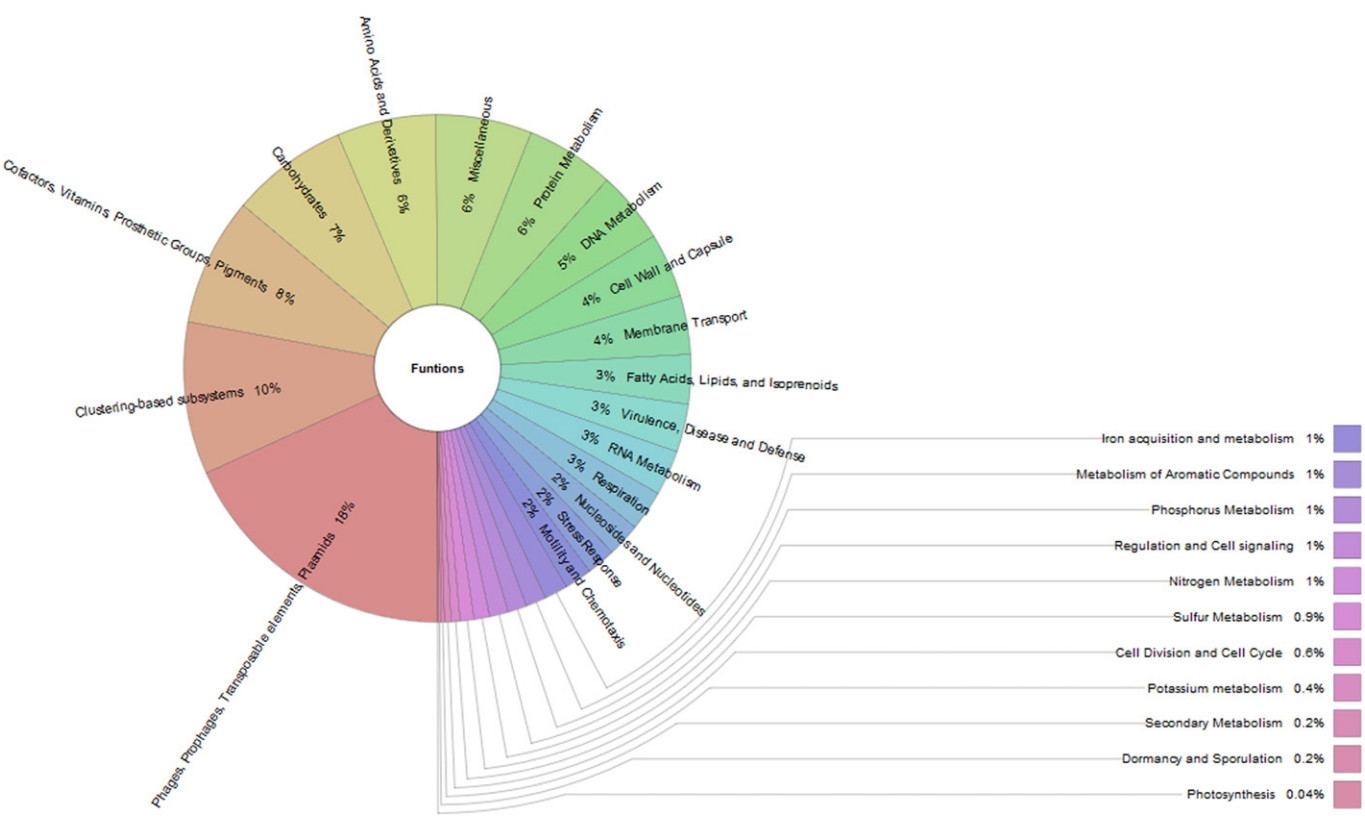


Fig. 2. Functional structure of Coamo thermal spring metagenome using subsystem annotation.

libraries by shotgun cloning over direct sequencing is that it allows for further functional tests from the initial isolate. Metagenomic sequencing from hot springs has led to the discovery of new phyla [5] and novel genes that encode for proteins with either known [2,6,11] and unknown functions [16]. Recently, thermostable enzymes such as the Fe-superoxide dismutase [9] and a DNA polymerase [14] have been identified using metagenomic analysis from thermal springs. These enzymes have the potential to be exploited for cosmetic and biotechnological purposes, respectively.

Puerto Rico has two principal hot springs in the southwest area. Coamo thermal spring (ThS) (18°02'16.6"N 66°22'27.6"W) is the warmest one. It is described as a sulfuric alkaline water with an average temperature of 43 °C [7]. Visitors from the local community and international origins often bathe in these waters due to its therapeutic cultural attributions. In Puerto Rico, the physicochemical and geological features of this hot spring are well characterized. To the best of our knowledge, its microbial diversity has not been studied before using a metagenomic approach. This project aimed to provide, for the first time, a community and functional microbial diversity profile of the water from Coamo ThS in Puerto Rico, using metagenomic library generation and shotgun sequencing.

## 2. Experimental design, materials and methods

### 2.1. Sampling

Water samples from the ThS located at Coamo, Puerto Rico (18°02'16.6"N 66°22'27.6"W) were collected at 0.3 meter depth using sterile 1-liter water sampling plastic containers and transported to the laboratory. pH and temperature values of the ThS were 8.22, and 47 °C, respectively.

### 2.2. Metagenomic library generation

For the metagenomic library generation, five (5) liters of ThS water samples were membrane-filtered using 0.45 µm MF-Millipore™. Membranes with trapped cells were slide-cut and prepared for DNA extraction using the Metagenomic DNA (metaDNA) Isolation kit from Water (Epicentre, USA) following the manufacturer's specifications. Extracted high molecular weight (40 kbp) DNA was end-repaired and ligated into the fosmid vector pCCFOS1. Recombinant DNA was packaged and transduced to *Escherichia coli* Epi300-T1<sub>R</sub> via T1 bacteriophages using CopyControl™ Fosmid Library Production Kit (Epicentre, USA) following the manufacturer's specifications. After plating the library, the total number of clones in the metagenomic library was determined. Random clones were selected to confirm the presence of pCCFOS1 vector and metaDNA by restriction enzyme digestion using *Bam*H1 (New England Biolabs).

### 2.3. Metagenome sequencing and pCCFOS1 vector removal

MetaDNA was extracted and purified using Midiprep kit (QIAGEN, USA) following manufacturer's specifications. It was sent to MR DNA (<http://www.mrdnlab.com>) where a genomic library was generated using Nextera DNA Sample Preparation Kit (Illumina) and Qubit® dsDNA HS Assay Kit (Life Technologies) following the manufacturer's specifications. After DNA samples were measured (50 ng) and diluted (to 2.5 ng/µL), fragmentation and addition of adapter sequences was done. The library was diluted (to 12 pM) and, then, sequenced using the 600 cycle v3 Reagent Kit (Illumina) on the MiSeq (Illumina). The pCCFOS1 vector was removed from the metagenomic library sequences using SeqClean (The Gene Indices Sequences Cleaning and Validation script) [15] to obtain solely the metagenome sequences.

### 2.4. Taxonomic and functional insights

To generate a taxonomic profile and a functional *in silico* description of the samples, the metagenomic sequencing data was analyzed with the Rapid Annotation using Subsystems Technology for metagenomes (MG-RAST) online server.

## 3. Results and discussion

The metagenomic library consisted of approximately 6000 clones and restriction enzyme digestion showed that 90% of the library contains metaDNA. After removing pCCFOS1, 122,026 sequences with 33.10 Mbps size with 64% of G + C content were annotated and analyzed using the MG-RAST online server [12]. Bacteria was the most abundant domain (95.84%) followed by unidentified sequences (2.28%), viruses (1.67%), eukaryotes (0.15%), and archaea (0.01%). The metagenome was constituted of a total of 48 microbial phyla from which the most abundant was *Proteobacteria* (95.03%), followed by unidentified (2.28%), unclassified from viruses (1.74%), *Firmicutes* (0.20%) and *Actinobacteria* (0.18%). In terms of bacterial families, the metagenome comprised of a total of 296 with an abundance distribution of *Enterobacteriaceae* (77.73%) followed by *Comamonadaceae* (6.36%), *Xanthomonadaceae* (2.95%) and unidentified families (2.28%). The most abundant species were *Escherichia coli*, *Polaromonas naphthalenivorans*, *Albidiferax ferrireducens* and *Acidovorax* sp. (Fig. 1). Moreover, *P. naphthalenivorans* produce dioxygenases capable of degrading naphthalene [10] and *A. ferrireducens* contains temperature resistant proteins that allowed it to grow under psychrophilic and thermophilic conditions [1].

Of the sequences that successfully passed the quality control tests (86.36% of the total) 79.50% contained predicted proteins with known functions, 19.95% had predicted proteins with unknown functions and 1% were ribosomal RNA genes. Functional analysis using Non-supervised Orthologous Groups (NOG) annotation showed that 82.79% of proteins are poorly characterized, indicating the possibility of potential reservoirs of novel microbial functions. Subsystem functional analysis (illustrated in Fig. 2) showed that 20% of genes belong to transposable elements (phages, prophages, and plasmids), 10% of genes belong to clustering-based subsystems, 8% of genes belong to the production of cofactors and secondary metabolites (vitamins, prosthetic groups and pigments). In addition to the previous analysis, genes belonging to metabolism of carbohydrates (7%), miscellaneous (6%), amino acids (6%), proteins (5%), DNA (4%), cell wall (4%) and membrane transport (3%) are also mentioned. Furthermore, the metagenome contains genes related to metabolism of iron, phosphorous and sulfur; analysis that provide insights of potential microbial roles in the ecology of biogeochemical cycles and ecosystem stability.

This study provides a comprehensive description and analysis of the microbial diversity and functional potential present in the metagenome of Coamo ThS. Using this data, now it is possible to design and develop feasible and pragmatic functional screenings of the metagenomic library generated for research and educational purposes.

### Nucleotide accession number

Metagenome sequence data are available on NCBI under the accession no. SAMN06131862.

### Acknowledgements

We would like to thank the University of Puerto Rico RISE2BEST Program (NIH-R25GM088023), the Puerto Rico Louis Stroke Alliance for Minority Participation (NSF-1400868) and Maximizing Access to Research Careers U\*STAR program (NIH-2T34GM008419-26) for supporting this project. In addition, we thank the administration of the Thermal Springs in Coamo, Puerto Rico for allowing us to collect

the samples. Finally, we mention and thank Hector M. Nieves-Rosado for his contribution in the manuscript generation.

## References

- [1] A.M. Achberger, B.C. Christner, A.B. Michaud, J.C. Prisco, M.L. Skidmore, T.J. Vick-Majors, the WISSARD Science Team, Microbial community structure of subglacial Lake Whillans, West Antarctica. *Front. Microbiol.* 7 (2016) 1457, <http://dx.doi.org/10.3389/fmicb.2016.01457>.
- [2] P. Brumm, M.L. Land, L.J. Hauser, C.D. Jeffries, Y.-J. Chang, D.a. Mead, Complete genome sequences of *Geobacillus* sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot spring in Yellowstone National Park. *Stand. Genomic Sci.* 10 (2015) 81, <http://dx.doi.org/10.1186/s40793-015-0075-0>.
- [3] S. Chan, K.G. Chan, Y.L. Tay, Y.H. Chua, K.M. Goh, Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* 6 (2015) 1–15, <http://dx.doi.org/10.3389/fmicb.2015.00177>.
- [4] D.R. Colman, Z.J. Jay, W.P. Inskeep, R.d.M. Jennings, K.R. Maas, D.B. Rusch, C.D. Takacs-Vesbach, Novel, deep-branching heterotrophic bacterial populations recovered from thermal spring metagenomes. *Front. Microbiol.* (2016) 1–14, <http://dx.doi.org/10.3389/fmicb.2016.00304>.
- [5] E.a. Eloë-Fadrosch, D. Paez-Espino, J. Jarett, P.F. Dunfield, B.P. Hedlund, A.E. Dekas, N.N. Ivanova, Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* 7 (2016) 10476, <http://dx.doi.org/10.1038/ncomms10476>.
- [6] R. Gupta, T. Govil, N. Capalash, P. Sharma, Characterization of a glycoside hydrolase family 1  $\beta$ -galactosidase from hot spring metagenome with transglycosylation activity. *Appl. Biochem. Biotechnol.* 168 (6) (2012) 1681–1693, <http://dx.doi.org/10.1007/s12010-012-9889-z>.
- [7] S. Guzman-Rios, Hydrology and water quality of the principal springs in Puerto Rico. Water-Resources Investigations Report from the U.S. Geological Survey, USGS Publications Warehouse, 1988 (<http://pubs.er.usgs.gov/publication/wri854269>).
- [8] J. Handelsman, Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* (2004) 669–685, <http://dx.doi.org/10.1128/MMBR.68.4.669-685.2004>.
- [9] Y.Z. He, K.Q. Fan, C.J. Jia, Z.J. Wang, W.B. Pan, L. Huang, Z.Y. Dong, Characterization of a hyperthermostable Fe-superoxide dismutase from hot spring. *Appl. Microbiol. Biotechnol.* 75 (2) (2007) 367–376, <http://dx.doi.org/10.1007/s00253-006-0834-3>.
- [10] C. Jeon, W. Park, P. Padmanabhan, C. DeRito, J. Snape, E. Madsen, Discovery of a novel bacterium with distinctive dioxygenase that is responsible for in situ biodegradation in contaminated sediment. *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 13591–13596.
- [11] O. López-López, K. Knapik, M.E. Cerdán, M.I. González-Siso, Metagenomics of an alkaline hot spring in Galicia (Spain): microbial diversity analysis and screening for novel lipolytic enzymes. *Front. Microbiol.* (2015) 1–18, <http://dx.doi.org/10.3389/fmicb.2015.01291>.
- [12] F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E.M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, R.A. Edwards, The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinforma.* 9 (2008) 386 (<http://www.biomedcentral.com/1471-2105/9/386>).
- [13] S.R. Miller, A.L. Strong, K.L. Jones, M.C. Ungerer, Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl. Environ. Microbiol.* 75 (13) (2009) 4565–4572, <http://dx.doi.org/10.1128/AEM.02792-08>.
- [14] M.J. Moser, R.A. DiFrancesco, K. Gowda, A.J. Klingele, D.R. Sugar, et al., Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS One* 7 (6) (2012), e38371. <http://dx.doi.org/10.1371/journal.pone.0038371>.
- [15] SeqClean, <https://sourceforge.net/projects/seqclean/>.
- [16] B. Szalkai, I. Scheer, K. Nagy, B.G. Vértessy, V. Grolmusz, The metagenomic telescope. *PLoS One* 9 (7) (2014)<http://dx.doi.org/10.1371/journal.pone.0101605>.