

Characterizing the microbial metagenome of calcareous stromatolite formations in the San Felipe Creek in the Anza Borrego Desert

Rosalina Stancheva,^{1,2} Arun Sethuraman,³ Hossein Khadivar,¹ Jenna Archambeau,^{1,4} Ella Caughran,^{1,5} Ashley Chang,^{1,6} Brad Hunter,¹ Christian Ihenyen,^{1,7} Marvin Onwukwe,^{1,8} Dariana Palacios,^{1,9} Chloe La Prairie,^{1,10} Nicole Read,¹ Julianna Tsang,^{1,11} Brianna Vega,¹ Cristina Velasquez,¹ Xiaoyu Zhang,¹ Elinne Becket,¹ Betsy Read¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We describe the metagenome composition, community functional annotation, and prokaryote diversity in calcareous stromatolites from a dry stream bed of the San Felipe Creek in the Anza Borrego Desert. Analyses show a community capable of nitrogen fixation, assimilatory nitrate reduction, biofilm formation, quorum sensing, and potential thick-walled akinete formation for desiccation resistance.

KEYWORDS stromatolite, metagenome

Stromatolites represent some of the earliest forms of marine cyanobacterial life, dating back 3.5 billion years (1, 2). Stromatolites are distributed from hypersaline coastal mats of the Hamelin Pools of Shark Bay in Western Australia (3) to oligotrophic lakes in the Chihuahuan Desert (4). Stromatolite freshwater habitats are extreme and exposed to desiccation events (5). We studied the temporally dry San Felipe Creek in the Anza Borrego Desert, Southern California, where stromatolites were previously identified (6).

Materials were collected on 20 November 2019 from granite rock tops in the San Felipe Creek (Fig. 1) (33.0986, -116.4708) (7). A single, dry stromatolite (2 cm thick) was chiseled, transported to California State University, San Marcos (CSUSM), and material (500 mg) from the entire sample, without targeting any specific layer, was ground with a bead mill. Cells were extracted with PEG-NaCl buffer prior to isolating DNA (8). Four replicate metagenomic libraries were prepared from a single stromatolite using the TELL-Seq WGS Library Prep Kit (Universal Sequencing, Carlsbad, CA, USA). Pooled libraries were sequenced on Illumina NextSeq 500/550 platform using the Mid Output Kit V2.5 (150 cycles, paired-end reads). Raw reads (FASTQ) were assessed using FastQC v.11.9 (Q score ≥ 30), and *de novo* metagenomes were assembled with UST TELL-Seq assembly pipeline. Metagenome quality was assessed with QUAST v.4.4 (8) on KBase v.2.6.4 (9). Kaiju v.1.7.3 (10) against the NCBI microbial genomes database in KBase v.2.6.4 (9) was used to classify operational taxonomic units (OTUs) from normalized raw reads (11). Automated gene calling was performed using NCBI PGAP (12) with *Lyngbya aestuarii* and *Sediminibacterium* as references. Predicted proteins were classified using GhostKOALA v.2.2 (13).

Pooled sequencing generated $\sim 2.05 \times 10^7$ reads with $>90.6\%$ exceeding $Q \geq 30$ and having a GC content of $\sim 54\%$ GC. While 2.8 Gb of sequence data was generated from each replicate library, the average metagenome assembly was 1.3 Mb (10 $\times$ depth). Optimized assembly with pooled reads yielded a larger, contiguous metagenome (~ 29 Mb). The largest scaffold was ~ 1.1 Mb with an L_{50} count of 109 and N_{50} size of 56 Kb (Table 1).

Editor Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Betsy Read, bread@csusm.edu.

The authors declare no conflict of interest.

See the funding table on p. 4.

Received 17 September 2023

Accepted 15 February 2024

Published 4 March 2024

Copyright © 2024 Stancheva et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

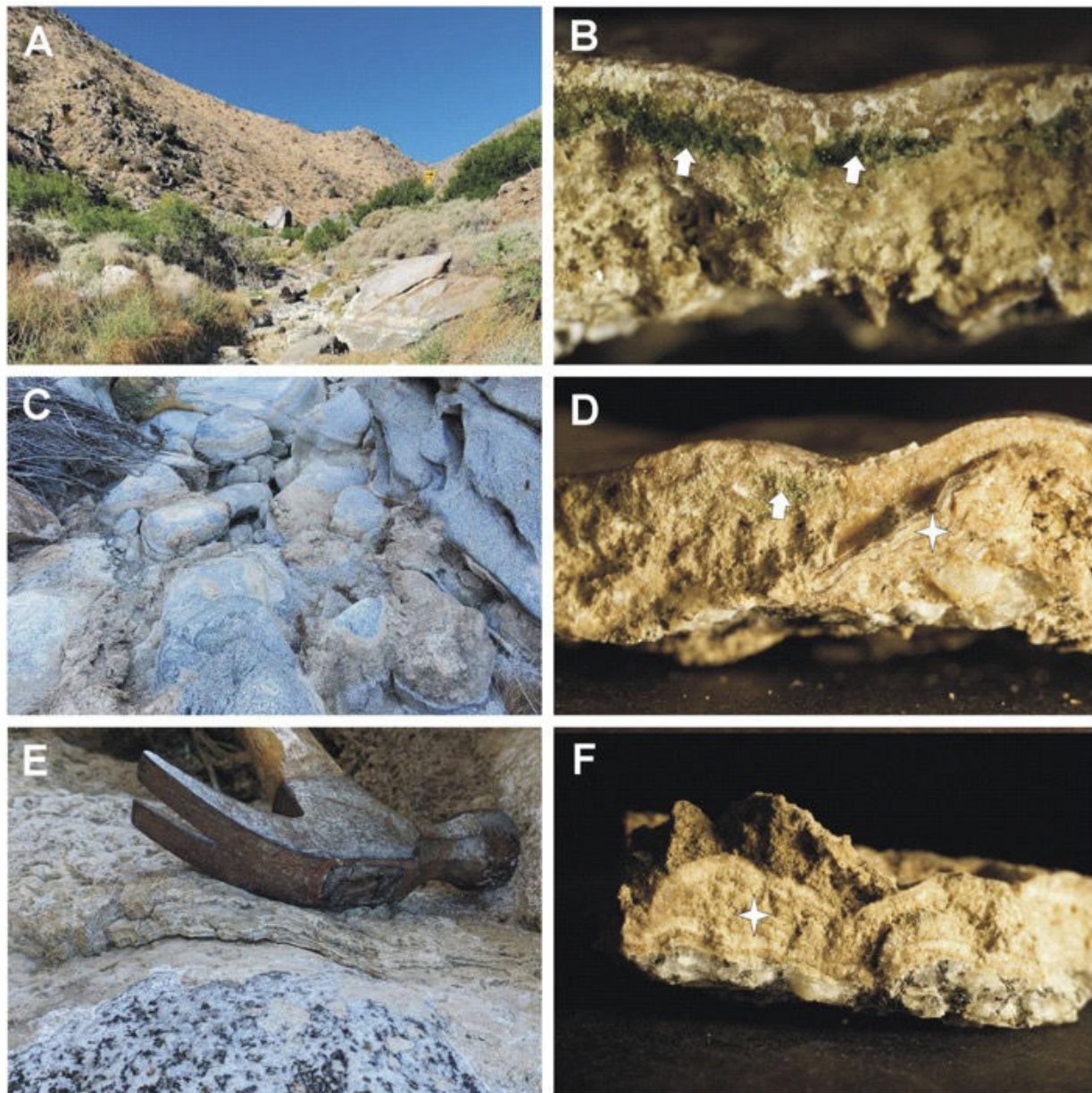


FIG 1 The ephemeral San Felipe Creek in Anza Borrego Desert (A). Samples were collected from the surface of boulders found in the bed of the desiccated creek (C and E). Samples imaged under a dissecting scope (B, D, and F) show characteristic features of stromatolites including endolithic cyanobacteria underneath a layer of calcite (arrows in panels B and D) and multiple laminate layers of organic and inorganic materials (asterisks in panels D and F). DNA was isolated from the stromatolites shown in panels B and E.

Classification of raw read OTUs showed that the prokaryotic communities were primarily composed of Cyanobacteria (85.95%, 45 genera), followed by Proteobacteria (5.73%) and Actinobacteria (4.33%). The taxonomic composition of Anza Borrego stromatolites was similar to freshwater stromatolites from pools in Cuatros Cienegas Basin (14) and Ruidera, Spain (5), which were also dominated by cyanobacteria (74% and 54%, respectively).

Gene prediction using PGAP resulted in community-level annotation of ~33,000 protein-coding genes (Table 1), including 7 complete 5S rRNA, 16 16S rRNA, and 34 23S partial rRNA genes. Of these, 11,732 (37%) were classified, 10% being affiliated with signaling and cellular processes like nitrogen cycling, quorum sensing, biofilm formation,

TABLE 1 Quality metrics were performed on the combined metagenomics data using QUAST and show a relatively large and contiguous 29 Mb metagenome

Parameter/statistic genome assembly	Metagenome
Number of scaffolds	5,478
Total scaffold size (Mb)	29.03
Longest scaffold (Mb)	1.13
Shortest scaffold (bp)	500
Number of scaffolds >1,000 nt	2,788 (50.9%)
Number of scaffolds >10,000 nt	355 (6.5%)
Number of scaffolds >100,000 nt	38 (0.7%)
Number of scaffolds >1,000,000 nt	1 (0.0%)
Number of scaffolds >10,000,000 nt	0 (0.0%)
Mean scaffold size (bp)	5,300
Median scaffold size (bp)	1,021
<i>N</i> ₅₀ scaffold length (bp)	56,004
<i>L</i> ₅₀ scaffold count	109
%GC	41.44
Genome annotation	
Genes (total)	33,494
Coding Sequences (CDS with protein homology)	33,203
Genes (RNA)	291
rRNAs	7, 16, 34 (5S, 16S, 23S)
tRNAs	221
ncRNAs	13
Pseudogenes (total)	1,127
CRISPR	65

and desiccation resistance. Additionally, several genes were linked to genetic information processing (9%), carbohydrate metabolism (9%), and environmental information processing (9%). Metagenomic signatures for morphological and physiological strategies to cope with the harsh conditions of the Anza Borrego Desert were found, including those for UV protective pigments scytonemin and carotenoids, and the synthesis of potential mycosporine-like amino acids; genes involved in microalgal desiccation tolerance, including those encoding aquaporins, chaperones, and antioxidants; and enzymes responsible for trehalose, sucrose, and polyamine synthesis.

ACKNOWLEDGMENTS

A.S. and B.R. thank the mentors of the Summer 2022 CSUSM NSF REU scholars for their continued support and encouragement for student-led research. We thank Douglas B. Read, Jr. for the stromatolite images presented in Fig. 1.

This work was funded by NSF-REU: 1852189 to PI B.R., co-PI A.S., and senior personnel R.S. and E.B., and was conducted over the Summer 2022 by CSUSM NSF REU students whose experiences were chronicled at <https://csusmbioreu.weebly.com/>. This research includes calculations carried out on HPC resources supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the U.S. Army Research Laboratory under contract number W911NF-16-2-0189.

AUTHOR AFFILIATIONS

- ¹Department of Biological Sciences, California State University San Marcos, San Marcos, California, USA
- ²Department of Environmental Science and Policy, George Mason University, Fairfax, Virginia, USA
- ³Department of Biology, San Diego State University, San Diego, California, USA

⁴Department of Chemistry, American University, Washington, DC, USA
⁵Department of Biological Sciences, University of New Hampshire, Durham, New Hampshire, USA
⁶Department of Mathematics, Amherst College, Amherst, Massachusetts, USA
⁷Department of Biology, Howard University, Washington DC, USA
⁸Department of Biological Sciences, University of Maryland, Baltimore County, Maryland, USA
⁹Department of Biosciences, Farmingdale State College, Farmingdale, New York, USA
¹⁰Department of Biology, Millikin University, Decatur, Illinois, USA
¹¹Department of Biological Sciences, Willamette University, Salem, Oregon, USA

AUTHOR ORCIDs

Rosalina Stancheva  <http://orcid.org/0000-0003-0562-4026>
Arun Sethuraman  <http://orcid.org/0000-0002-8201-8292>
Hossein Khadivar  <http://orcid.org/0000-0001-6673-8540>
Marvin Onwukwe  <http://orcid.org/0000-0003-1428-4983>
Julianna Tsang  <http://orcid.org/0000-0003-0835-4148>
Elinne Becket  <http://orcid.org/0000-0002-2504-024X>
Betsy Read  <http://orcid.org/0000-0001-7258-8032>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	1852189	Betsy Read Arun Sethuraman

AUTHOR CONTRIBUTIONS

Rosalina Stancheva, Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review and editing | Arun Sethuraman, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review and editing | Hossein Khadivar, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Ella Caughran, Methodology | Ashley Chang, Methodology | Brad Hunter, Methodology | Christian Ihenyen, Methodology | Marvin Onwukwe, Methodology | Dariana Palacios, Methodology | Chloe La Prairie, Methodology | Nicole Read, Investigation, Methodology | Julianna Tsang, Methodology | Brianna Vega, Investigation | Cristina Velasquez, Methodology | Xiaoyu Zhang, Supervision, Visualization | Elinne Becket, Data curation, Investigation, Methodology, Supervision | Betsy Read, Conceptualization, Data curation, Formal analysis, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing.

DATA AVAILABILITY

The assembled metagenome has been submitted to NCBI and is accessible via BioProject: [PRJNA967693](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA967693). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession [JASEJY010000000](https://www.ncbi.nlm.nih.gov/seq/assembly/JASEJY010000000). The version described in this paper is version [JASEJY010000000](https://www.ncbi.nlm.nih.gov/seq/assembly/JASEJY010000000). Scripts and code for all analyses performed can be accessed via the project's GitHub page, accessible at https://github.com/j-archambeau/Stromatolite_Metagenomics

REFERENCES

- Schopf JW. 2006. Fossil evidence of Archaean life. *Philos Trans R Soc Lond B Biol Sci* 361:869–885. <https://doi.org/10.1098/rstb.2006.1834>
- Schopf JW, Packer BM. 1987. Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona group, Australia. *Science* 237:70–73. <https://doi.org/10.1126/science.11539686>

3. Playford PE, Cockbain AE. 1976. Modern algal stromatolites at hamelin pool, a hypersaline barred basin in shark Bay, Western Australia, p 389–411. In Walter MR (ed), *Stromatolites - developments in sedimentology*. Amsterdam: Elsevier.
4. Falcón LI, Cerritos R, Eguarte LE, Souza V. 2007. Nitrogen fixation in microbial mat and stromatolite communities from cuatro Cienegas, Mexico. *Microb Ecol* 54:363–373. <https://doi.org/10.1007/s00248-007-9240-3>
5. Santos F, Peña A, Nogales B, Soria-Soria E, Del Cura MAG, González-Martín JA, Antón J. 2010. Bacterial diversity in dry modern freshwater stromatolites from ruidera pools natural Park, Spain. *Syst Appl Microbiol* 33:209–221. <https://doi.org/10.1016/j.syapm.2010.02.006>
6. Buchheim P. 1995. Stromatolites: living fossils in ANZA-borrego desert state park. In Paul Remeika, Anne Sturz (ed), *Field trip guidebook and volume for the 1995 San Diego Association of geologist's field trip to ANZA-borrego desert state park*. San Diego Association of Geologists.
7. Narayan A, Jain K, Shah AR, Madamwar D. 2016. An efficient and cost-effective method for DNA extraction from athalassohaline soil using a newly formulated cell extraction buffer. *3 Biotech* 6:62. <https://doi.org/10.1007/s13205-016-0383-0>
8. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
9. Allen B, Drake M, Harris N, Sullivan T. 2017. Using KBase to assemble and annotate prokaryotic genomes. *Curr Protoc Microbiol* 46:1E. <https://doi.org/10.1002/cpmc.37>
10. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257. <https://doi.org/10.1038/ncomms11257>
11. Beule L, Karlovsky P. 2020. Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. *PeerJ* 8:e9593. <https://doi.org/10.7717/peerj.9593>
12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
13. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
14. Breitbart M, Hoare A, Nitti A, Siefert J, Haynes M, Dinsdale E, Edwards R, Souza V, Rohwer F, Hollander D. 2009. Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. *Environ Microbiol* 11:16–34. <https://doi.org/10.1111/j.1462-2920.2008.01725.x>