Title: Whole-Genome Sequencing of Three Extremophile *Bacillus subtilis Strains* Isolated from the Atacama Desert

Authors: Nicole T Cavanaugh[1], Girish Kumar[2], Matteo Couto Frignani[3], Nathan Thewedros[1], Marcello Twahirwa[1], Carlos Riquelme[4], André O Hudson[2], Yunrong Chai[1], Veronica Godoy-Carter[1*]

Affiliations:
[1]Northeastern University College of Science, Department of Biology
[2]Rochester Institute of Technology College of Science, Gosnell School of Life Sciences
[3]Northeastern University College of Engineering, Department of Bioengineering
[4]University of Antofagasta, Department of Biotechnology
[*]Corresponding Author: v.godoycarter@northeastern.edu

Abstract:
The Atacama Desert is home to many bacteria that use biofilm formation to protect themselves against the harsh environment. We isolated and sequenced three soil *Bacillus subtilis* isolates that were isolated from Atacama soil. These sequences will aid research in survival mechanisms of eubacteria in the Atacama Desert.

Announcement:
The Atacama Desert is one of the most ancient with the highest radiation and driest desert in the world (1). Despite this, it is home to a variety of bacterial life. Bacteria in this region are shown to have developed adaptations to survive this harsh, irradiated environment such as high levels of pigment production and biofilm formation (2). In the past, most research addressing survival mechanisms in microbes from the Atacama have focused on archaea. We are studying how eubacteria have developed these adaptations to survive in a desert environment.

To study how biofilm formation impacts bacterial survival in the area, we started by culturing and further isolating a variety of bacterial colonies from the desert soil and screened for biofilm-forming isolates. Bacterial colonies were purified after plating 100 µL of a soil water slurry on Reasoner's 2A (R2A) agar medium and incubating plates at 28 ºC for 48-72 hours. To identify the isolates, we performed 16S rRNA gene sequencing on the V3/4 variable regions using the following primers: 5'-CCTACGGGNGGCWGCAG-3' and 5'-GACTACHVGGGTATCTAATCC-3'. Using National Society for Biotechnology Information (NCBI) BLASTN version 2.15.0, three *Bacillus sp*. were identified (3, 4). The *Bacillus* genus contains Gram-positive bacteria that are found throughout the environment, many of which are known to produce biofilms (5, 6, 7, 8, 9, 10). To test biofilm formation, strains were grown at 37 ºC to OD600 = 1.0 in Luria-Bertani (LB) broth and 2 µl of culture was spotted on LB agar plates supplemented with 1% glycerol and 0.1 mM $MnSO_4$ (LBGM, biofilm inducing medium) (11). Colony biofilms were incubated for 48 hours.

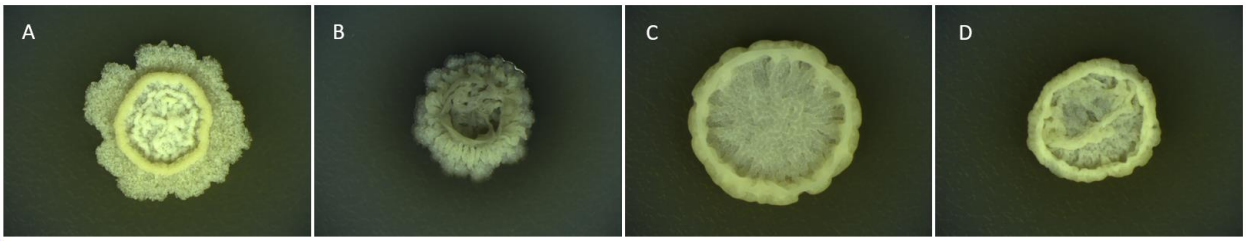


Figure 1: Colony biofilms formed by environmental strains from the Atacama Desert*. Colony biofilms of 0102A (A), 0209A (B), 0909A (C), and 3610 (*B. subtilis *lab strain) (D) grown on LBGM medium after 48 hours of incubation.*

These *Bacillus sp*. were selected for further genomic analysis. Genomic DNA was isolated from a 2-mL LB culture using the Wizard Genomic DNA Purification Kit (Promega, USA) following kit instructions for Gram-positive bacteria. Purified DNA was quantified using a NanoDrop spectrophotometer (Thermo-Fisher), DNA libraries were prepared using the Nextera XT library preparation kit (Illumina) and sequenced using the Illumina MiSeq System at the Genomics Lab, Rochester Institute of Technology. Raw reads were trimmed using Trimmomatic Galaxy version 0.38.1 (12). The trimmed reads were assembled de novo using Unicycler Galaxy version 0.5.0+galaxy1 (13). Quality analysis was performed using Quast Galaxy version 5.2.0+galaxy1 (14) and FastQC Galaxy version 0.74+galaxy0 (15). Quality assessments are outlined in Table 1. All programs used above were run at default parameters unless otherwise noted. Further analysis of each genome using NCBI BLASTN Galaxy Version 2.14.1+galaxy showed that all three sequences were *Bacillus subtilis* (3-4), an environmental bacterium commonly found in the soil known to form beneficial biofilm on plant roots (5, 6, 10).

Isolate	Total Bases (Mbp)	Read Length	Number of Reads (FastQC)	Number of Contigs	N50	GC Content
0102A	813.2	4,192,438 bp	4,406,089	27	350,970 bp	43.87%
0209A	980.4	4,167,904 bp	5,239,849	19	1,048,338 bp	43.98%
0909A	857.1	3,899,559 bp	4,586,376	19	930,776 bp	43.66%

Table 1: Summary of quality assessments for each strain. *A summary of the identities and quality assessments of each strain 0102A, 0209A, and 0909A. For each strain, the total bases and number of reads was determined using FastQC Galaxy version 0.74+galaxy0. Read length, number of contigs, N50, and GC content were assessed using Quast Galaxy version 5.2.0+galaxy1.*

Data Availability:

The WGS projects for *Bacillus subtilis* strains 0102A, 0209A, and 0909A have been deposited in GenBank under BioProject number PRJNA1071648. Strain 0102A is deposited under BioSample number SAMN39706723 and SRA number SRR27839954. Strain 0209A is

deposited under BioSample number SAMN39706724 and SRA number SRR27839953. Strain 0909A is deposited under BioSample number SAMN39706725 and SRA number SRR27839952.

The NCBI BLASTN webserver was used to identify the 0102A, 0209A, and 0909A 16s rRNA gene sequence. All programs found on the Galaxy platform can be found at usegalaxy.org.

Acknowledgements:

We would like to thank the National Science Foundation (MCB1651732) and the NU Global Experience Office for their support. Thank you to the RIT Genomics Lab. V.G. was supported by the HHMI Inclusive Excellence grant and the Northeastern Global Experiential Office. M.C.F, N.T., and M.T. were supported by NU Undergraduate Research and Fellowships Office PEAK Awards. N.C. is supported by the NSF Graduate Research Fellowship Program (1938052).

References:

[1] Bull, A. T., Andrews, B. A., Dorador, C., & Goodfellow, M. (2018). Introducing the Atacama Desert. *Antonie van Leeuwenhoek*, *111*(8), 1269–1272. https://doi.org/10.1007/s10482-018-1100-2

[2] Reverdy, A., Hathaway, D., Jha, J., Michaels, G., Sullivan, J., Mac-Adoo, D. D., Riquelme, C., Chai, Y., & Godoy, V. G. (2020). *Insights into the diversity and survival strategies of soil bacterial isolates from the Atacama Desert* (p. 2020.09.24.312199). bioRxiv. https://doi.org/10.1101/2020.09.24.312199

[3] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

[4] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

[5] Arnaouteli, S., Bamford, N. C., Stanley-Wall, N. R., & Kovács, Á. T. (2021). Bacillus subtilis biofilm formation and social interactions. *Nature Reviews Microbiology*, *19*(9), Article 9. https://doi.org/10.1038/s41579-021-00540-9

[6] Beauregard, P. B., Chai, Y., Vlamakis, H., Losick, R., & Kolter, R. (2013). Bacillus subtilis biofilm induction by plant polysaccharides. *Proceedings of the National Academy of Sciences*, *110*(17), E1621–E1630. https://doi.org/10.1073/pnas.1218984110

[7] Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R., & Kolter, R. (2001). Fruiting body formation by Bacillus subtilis. *Proceedings of the National Academy of Sciences*, *98*(20), 11621–11626. https://doi.org/10.1073/pnas.191384198

[8] Diomande, S., NGUYEN-THE, C., Guinebretière, M.-H., Broussolle, V., & Brillard, J. (2015). Role of fatty acids in Bacillus environmental adaptation. *Frontiers in Microbiology*, *6*. https://www.frontiersin.org/articles/10.3389/fmicb.2015.00813

[9] Fritze, D. (2004). Taxonomy of the Genus Bacillus and Related Genera: The Aerobic Endospore-Forming Bacteria. *Phytopathology®*, *94*(11), 1245–1248. https://doi.org/10.1094/PHYTO.2004.94.11.1245

[10] Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., & Kolter, R. (2013). Sticking together: Building a biofilm the Bacillus subtilis way. *Nature Reviews Microbiology*, *11*(3), Article 3. https://doi.org/10.1038/nrmicro2960

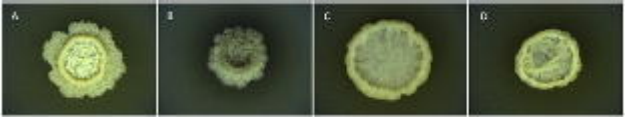
[11] Shemesh, M., & Chai, Y. (2013). A Combination of Glycerol and Manganese Promotes Biofilm Formation in Bacillus subtilis via Histidine Kinase KinD Signaling. *Journal of Bacteriology*, *195*(12), 2747–2754. https://doi.org/10.1128/JB.00028-13

[12] Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

[13] Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, *13*(6), e1005595. https://doi.org/10.1371/journal.pcbi.1005595

[14] Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266

[15] *Babraham Bioinformatics—FastQC A Quality Control tool for High Throughput Sequence Data*. (n.d.). Retrieved February 2, 2024, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/


A
B
C
D