

# Draft genome sequence of *Geobacillus* sp. strain G4 isolated from the Calientes geothermal field in Tacna, Peru

Roberto Castellanos,<sup>1</sup> Karita C. R. Santos,<sup>2</sup> Alonso R. Poma Ticona,<sup>1</sup> Heber E. Ramirez-Arua,<sup>3</sup> Soledad A. Bornás-Acosta,<sup>1</sup> Jéssica Pinheiro Silva,<sup>4</sup> Pedro R. Vieira Hamann,<sup>4</sup> Fabyano A. C. Lopes<sup>2</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** We report the draft genome sequence of *Geobacillus* sp. strain G4, isolated from a hot spring in the Calientes geothermal field, Tacna, Peru. The genome was assembled into 164 contigs, totaling 3,362,411 bp, with an N50 value of 236,119 bp and a guanine-cytosine (GC) content of 52.59%.

**KEYWORDS** *Geobacillus*, thermophilic bacteria, geothermal spring, Calientes geothermal field

**G**eobacillus sp. strain G4, a thermophilic bacterium, was isolated from sediment samples (upper 15 cm layer) collected from a geothermal spring in the Calientes geothermal field, Candarave, Tacna, Peru (UTM 0.379908 E, 81.09185 N). A total of 250 g of sediment was collected in sterile containers, and 10 g was enriched and then plated on Luria-Bertani (LB) medium at 60°C for 48 hours (1). Strain G4 was isolated and confirmed as axenic after three transfers. Strain G4 showed high amylolytic activity, determined by the Miller method (2): 100 µL of starch (0.5% wt/vol) and 100 µL of G4 culture supernatant were incubated at 60°C for 15 minutes; 300 µL of DNS reagent was added, and absorbance was read at 540 nm.

*Geobacillus* sp. strain G4 was grown in 10 mL of LB medium at 60°C for 48 hours (1). Cells were harvested by centrifugation at 8,000 × g for 15 minutes at 4°C. Genomic DNA was isolated using the Bacterial Genomic DNA Isolation Kit (Norgen Biotek Corp., Canada) according to the manufacturer's instructions. The DNA libraries were prepared using the Illumina TruSeq Nano DNA Library Prep Kit (350), and sequencing was performed using an Illumina NovaSeq platform with paired-end reads, each 150 bases in length, by Macrogen Inc. (South Korea). The raw fastq files from the genome, consisting of 13,778,078 reads, were imported into the FastQC v0.12.1 tool for quality assessment (3). Quality filtering was performed using Trimmomatic v0.38. The remaining sequences were assembled using SPAdes v3.15.4. Short contigs (<200 bp) were removed from the assembly by seqtk v1.4 (<https://github.com/lh3/seqtk>). Assembly size and quality were evaluated using Quast v5.2.0 (4) and Busco v5.5.0 (5). Completeness and contamination parameters of the assembled genome were evaluated using CheckM v1.2.2 tool (6). NCBI Prokaryotic Genome Annotation Pipeline was used to annotate the genome using the RefSeq database (7). The 16S rRNA gene sequence of *Geobacillus* sp. strain G4 was extracted using barrnap v0.9 (<https://github.com/tseemann/barrnap>). Finally, for the identification of biosynthetic gene clusters related to secondary metabolism (BGCs) related to secondary metabolism in microorganisms, the AntiSMASH v7.1.0 software was utilized (8). Default parameters were used for all software tools unless otherwise noted.

The draft genome of *Geobacillus* sp. strain G4 was approximately 3.4 Mb with ~1,225x coverage. The assembly comprised 164 contigs with a G + C content of 52.59%. Genome completeness and contamination were 99.45% and 0.95%, respectively. The N50 and L50 values were 236,119 bp and 5 contigs, respectively. The annotation identified 3,490

**Editor** Julia A. Maresca, SUNY College of Environmental Science and Forestry, Syracuse, New York, USA

Address correspondence to Alonso R. Poma Ticona, [arpomat@unjbg.edu.pe](mailto:arpomat@unjbg.edu.pe).

Roberto Castellanos and Karita C. R. Santos contributed equally to this article. Author order was determined on the basis of seniority.

The authors declare no conflict of interest.

See the funding table on p. 2.

**Received** 17 July 2024

**Accepted** 8 November 2024

**Published** 21 November 2024

Copyright © 2024 Castellanos 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/).

genes: 2,047 coding sequences (CDS), 5 rRNAs, 88 tRNAs, 1 transfer-messenger RNA, and 1,349 putative proteins. Regarding gene clusters related to secondary metabolism (BGCs), five clusters were identified: betalactone, type III polyketide synthases, lanthipeptide-class-I, NI-siderophore, and terpene. The 16S rRNA gene sequence of *Geobacillus* sp. strain G4 showed a 99.87% identity with that of *Geobacillus kaustophilus* strain BGSC 90A1 (GenBank accession no. NR\_115285.2) and a 99.80% identity with *Geobacillus thermoleovorans* strain LEH-1 (GenBank accession no. NR\_036985.1). This genome enhances our understanding of thermophilic bacteria and their biotechnological potential.

## ACKNOWLEDGMENTS

This research was funded by the Jorge Basadre Grohmann National University through the "Canon, Sobrecano, and Mining Royalties" fund, as approved by Rectoral Resolution no. 7424-2020-UN/JBG, under the project "Application of Microbial Thermoenzymes for the Biodegradation of the Plant Organic Fraction of Municipal Solid Waste."

Alonso R. Poma Ticona acknowledges the financial support of CONCYTEC through PROCIENCIA Program, within the framework of Contest E067-2022-042 (contract no. PE501080334-2022-PROCIENCIA). Kárita C. R. Santos was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Fabyano A. C. Lopes was supported by Fundação de Amparo à Pesquisa do Tocantins (FAPT).

## AUTHOR AFFILIATIONS

<sup>1</sup>Enzyme Biotechnology Research Laboratory, Science Faculty, Universidad Nacional Jorge Basadre Grohmann, Tacna, Peru

<sup>2</sup>Laboratório de Microbiologia, Universidade Federal Do Tocantins, Porto Nacional, Brazil

<sup>3</sup>Laboratorio de Biología Molecular, Facultad de Farmacia y Bioquímica, Universidad Nacional Mayor de San Marcos, Lima, Peru

<sup>4</sup>São Carlos Institute of Physics, University of São Paulo, São Carlos, Brazil

## AUTHOR ORCIDs

Roberto Castellanos  <http://orcid.org/0000-0002-9021-412X>

Alonso R. Poma Ticona  <http://orcid.org/0000-0002-9027-1589>

Fabyano A. C. Lopes  <http://orcid.org/0000-0002-4565-9103>

## FUNDING

| Funder  | Grant(s)                                | Author(s)             |
|---|---|-----------------------|
| Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) | PE501080334-2022-PRO- CIENCIA           | Alonso R. Poma Ticona |
| Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)         |   | Karita C. R. Santos   |
| Jorge Basadre Grohmann National University                                  | Rectoral Resolution N° 7424-2020-UN/JBG | Roberto Castellanos   |

## AUTHOR CONTRIBUTIONS

Roberto Castellanos, Funding acquisition, Investigation, Supervision, Methodology | Karita C. R. Santos, Formal analysis, Investigation, Methodology | Alonso R. Poma Ticona, Funding acquisition, Supervision, Writing – review and editing | Heber E. Ramirez-Arua, Investigation, Methodology | Soledad A. Bornás-Acosta, Investigation | Jéssica Pinheiro Silva, Investigation | Pedro R. Vieira Hamann, Investigation | Fabyano A. C. Lopes, Investigation, Supervision, Writing – review and editing, Methodology

## DATA AVAILABILITY

The Illumina raw reads were uploaded to the SRA database under the accession number [SRR29836749](#). Genome assembly is available under the accession number [ASM4032797v1](#). The draft genome sequence of *Geobacillus* sp. strain G4 has been deposited at GenBank under the accession number [JBEPJY000000000](#), BioProject number [PRJNA1124973](#), and BioSample number [SAMN41878472](#).

## REFERENCES

1. Castañeda-Barreto A, Olivera-Gonzales P, Tamariz-Angeles C. 2024. A natural consortium of thermophilic bacteria from Huancarhuaz hot spring (Ancash-Peru) for promising lignocellulose bioconversion. *Helion* 10:e27272. <https://doi.org/10.1016/j.heliyon.2024.e27272>
2. Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* 31:426–428. <https://doi.org/10.1021/ac60147a030>
3. Andrews S. Babraham Bioinformatics. 2010. P. 1 FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
4. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
5. Seppey M, Manni M, Zdobnov EM. 2019. *Busco*: assessing genome assembly and annotation completeness. *Methods Mol Biol* 1962:227–245. [https://doi.org/10.1007/978-1-4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14)
6. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
7. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
8. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* 51:W46–W50. <https://doi.org/10.1093/nar/gkad344>