




Draft Genome Sequence of *Bacillus paralicheniformis* F47, Isolated from an Algerian Salty Lake

Mohamed Seghir Daas,^{a,b}  Albert Remus R. Rosana,^c Jeella Z. Acedo,^c Malika Douzane,^b Farida Nateche,^d Salima Kebbouche-Gana,^a John C. Vederas^c

^aValcore Laboratory, Department of Biology, University M'Hamed Bougara of Boumerdes, Boumerdes, Algeria

^bFood Technology Research Division, Institut National de la Recherche Agronomique d'Algérie, El Harrach, Algiers, Algeria

^cDepartment of Chemistry, University of Alberta, Edmonton, Alberta, Canada

^dMicrobiology Group, Laboratory of Cellular and Molecular Biology, Faculty of Biological Sciences, University of Science and Technology–Houari Boumediene, Bab Ezzouar, Algiers, Algeria

ABSTRACT *Bacillus paralicheniformis* F47 was isolated from a salty lake in Ain Baida-Ouargla, southern Algeria. The genome contains genes for the production of several bioactive secondary metabolites, including the siderophore bacillibactin, the lipopeptides fengycin, surfactin, and lichenysin, the antibiotics bacitracin and kanosamine, and a putative circular bacteriocin.

Rhizosphere-associated bacteria such as *Bacillus* spp. are able to suppress a variety of plant pathogens by producing antimicrobial secondary metabolites. For instance, *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* F11 can inhibit several Gram-positive and Gram-negative bacteria and fungal pathogens due to the production of antimicrobial compounds, including fengycin lipopeptides (1). Members of the *B. cereus sensu lato* group also produce a variety of bioactive small molecules such as lantibiotics and lipopeptides (2). Here, we report the draft genome of *Bacillus paralicheniformis* F47, isolated from a salty lake in southern Algeria, which contains multiple lipopeptides, siderophores, and a putative bacteriocin.

Genomic DNA was extracted using a DNeasy blood and tissue kit (Qiagen), as described previously (2). Sequencing libraries were created using the Nextera XT DNA library preparation kit, and sequencing was performed using the MiSeq reagent kit v2 (Illumina). *De novo* assembly of the 4,028,320-bp paired-end sequences (3.59 million reads) was generated using CLC Genomics Workbench v7.5.2. (CLC bio, Aarhus, Denmark), which resulted in 119 scaffolds, providing a coverage of 167× and G+C content of 46.2%. Gene prediction was performed using three automated annotation pipelines, Rapid Annotations using Subsystems Technology (RAST) v2.0 (3), the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (4), and the Joint Genome Institute-Integrated Microbial Genomes and Microbiomes (JGI-IMG/M) pipeline (5). Species identity was determined by calculating the average nucleotide identity (ANI) and alignment fraction (AF) using the Microbial Species Identifier (MiSI) calculator employed in the IMG/M system (6) against previously sequenced genomes in the IMG/M database. Strain identity was established by calculating the digital DNA:DNA hybridization (dDDH) using the Genome-to-Genome Distance Calculator (GGDC) 2.1 server (7). Secondary metabolites were predicted using antiSMASH v4 (8), and bacteriocins were predicted using BAGEL4 (<http://bagel4.molgenrug.nl>) (9). Genomic islands were predicted by using IslandViewer 4.0 (10), prophages by using PHASTER (11), and clustered regularly interspaced short palindromic repeats (CRISPRs) by using CRISPR-finder (12).

Received 14 February 2018 Accepted 28 February 2018 Published 29 March 2018

Citation Daas MS, Rosana ARR, Acedo JZ, Douzane M, Nateche F, Kebbouche-Gana S, Vederas JC. 2018. Draft genome sequence of *Bacillus paralicheniformis* F47, isolated from an Algerian salty lake. Genome Announc 6: e00190-18. <https://doi.org/10.1128/genomeA.00190-18>.

Copyright © 2018 Daas 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/).

Address correspondence to John C. Vederas, john.vederas@ualberta.ca.

M.S.D. and A.R.R.R. contributed equally to this article.

Genome annotation by the NCBI PGAP predicted 4,117 genes, including 4,026 coding sequences (CDSs), while RNA genes were composed of 11 rRNAs and 75 tRNAs. The genome of *B. paralicheniformis* F47 exhibits 99.26% ANI, 0.94 AF, and 70.54% dDDH with the type strain *B. paralicheniformis* KJ-16^T (13), supporting the placement of F47 in the species *B. paralicheniformis*. Conversely, the closest non-*B. paralicheniformis* species, *Bacillus licheniformis* ATCC 14580^T, exhibits 94.69% ANI, 0.88 AF, and 10.96% dDDH. Two putative CRISPR modules, one prophage, and three genomic islands were discovered in the genome.

The F47 genome contains several gene clusters with homology to known lipopeptide biosynthetic operons, including fengycin (*fenABCDE*) (1), surfactin (*srfADCB*) (14), and lichenysin (*lchAA* and *lchAB*) (15). The operons for the antibiotics bacitracin (scaffolds 11, 43, and 103) and kanosamine (*kabABC*) (16) were detected in the genome. Two siderophore gene clusters were also predicted in the genome, including bacillibactin and a putative anthrachelin (17). Lastly, a gene cluster of a putative circular bacteriocin was identified.

Accession number(s). This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank and JGI-IMG/M under the accession numbers [MYFI00000000](https://doi.org/10.1093/genomeA.00383-17) and Ga0181491, respectively.

ACKNOWLEDGMENTS

We gratefully acknowledge the help of Georgina Mcintyre from the Applied Genomic Core, University of Alberta, for assisting in library preparation and whole-genome sequencing.

This work was performed under the auspices of the Natural Sciences and Engineering Research Council of Canada to J.C.V. J.Z.A. was funded by Alberta Innovates–Health Solutions. This work, including the efforts of A.R.R.R., was funded by the Vanier Canada Graduate Scholarship, Alberta Innovates–Technology Futures, and the President’s Doctoral Prize of Distinction. M.S.D. was funded by Agence Universitaire de la Francophonie, France. M.S.D., M.D., F.N., and S.K.-G. were supported by the National Fund for Scientific Research of Algeria.

REFERENCES

- Daas MS, Acedo JZ, Rosana AR, Orata FD, Reiz B, Zheng J, Nateche F, Case RJ, Kebbouche-Gana S, Vederas JC. 2018. *Bacillus amyloliquefaciens* ssp. *plantarum* F11 isolated from Algerian salty lake as a source of biosurfactants and bioactive lipopeptides. *FEMS Microbiol Lett* 365: fnx248. <https://doi.org/10.1093/femsle/fnx248>.
- Daas MS, Rosana AR, Acedo JZ, Nateche F, Kebbouche-Gana S, Vederas JC, Case RJ. 2017. Draft genome sequences of *Bacillus cereus* E41 and *Bacillus anthracis* F34 isolated from Algerian salt lakes. *Genome Announc* 5:e00383-17. <https://doi.org/10.1128/genomeA.00383-17>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
- 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>.
- Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpidis NC. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* 42:D560–D567. <https://doi.org/10.1093/nar/gkt963>.
- Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavromatis K, Kyrpidis NC, Pati A. 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 43:6761–6771. <https://doi.org/10.1093/nar/gkv657>.
- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. <https://doi.org/10.1186/1471-2105-14-60>.
- Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Duran S, Hernandez G, de los Santos EL, Kim HU, Nave M. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
- van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP. 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides. *Nucleic Acids Res* 41:W448–W453. <https://doi.org/10.1093/nar/gkt391>.
- Bertelli C, Laird MR, Williams KP; Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FS. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 45:W30–W35. <https://doi.org/10.1093/nar/gkx343>.
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <https://doi.org/10.1093/nar/gkm360>.
- Dunlap CA, Kwon SW, Rooney AP, Kim SJ. 2015. *Bacillus paralicheniformis* sp. nov., isolated from fermented soybean paste. *Int J Syst Evol Microbiol* 65:3487–3492. <https://doi.org/10.1099/ijsem.0.000441>.
- Koumoutsis A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R. 2004. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J Bacteriol* 186: 1084–1096. <https://doi.org/10.1128/JB.186.4.1084-1096.2004>.
- Veith B, Herzberg C, Steckel S, Feesche J, Maurer KH, Ehrenreich P,

- Bäumer S, Henne A, Liesegang H, Merkl R, Ehrenreich A, Gottschalk G. 2004. The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. *J Mol Microbiol Biotechnol* 7:204–211. <https://doi.org/10.1159/000079829>.
16. Vetter ND, Langill DM, Anjum S, Boisvert-Martel J, Jagdhane RC, Omene E, Zheng H, van Straaten KE, Asiamah I, Krol ES, Sanders DA, Palmer DR. 2013. A previously unrecognized kanosamine biosynthesis pathway in *Bacillus subtilis*. *J Am Chem Soc* 135:5970–5973. <https://doi.org/10.1021/ja4010255>.
17. Challis GL. 2005. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chem Bio Chem* 6:601–611. <https://doi.org/10.1002/cbic.200400283>.