



Draft Genome Sequence of the Halophilic Strain *Citrobacter braakii* AN-PRR1, Isolated from Rhizospheric Soil of Rice (*Oryza sativa* L.) from Pakistan

Aniqa Nawaz,^{a,b} Fathia Mubeen,^b Zia Ul Qamar,^c Muhammad Usama Marghoob,^b Saefuddin Aziz,^{a,d}  Harald Gross^a

^aDepartment of Pharmaceutical Biology, Pharmaceutical Institute, University of Tübingen, Tübingen, Germany

^bMicrobial Physiology Lab, Soil and Environmental Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, Constituent College of Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan

^cDepartment of Nuclear Institute of Agriculture and Biology, Constituent College of Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan

^dMicrobiology Department, Biology Faculty, Jenderal Soedirman University, Purwokerto, Indonesia

ABSTRACT *Citrobacter braakii* AN-PRR1 is a potential salt-tolerant, plant growth-promoting rice rhizobacterium isolated from Pakistani soil. The 4.9-Mb draft genome sequence contributes to its taxonomic classification and will reveal the genes putatively responsible for its osmoprotectant and plant growth-promoting activity.

Rice, an important cereal worldwide, is under major threat from growing issues like salinity and drought due to the ongoing climate change (1–4). To combat the adverse effects of salinity on the production of rice and other crops, the use of salt-tolerant plant growth-promoting rhizobacteria (PGPR) is a promising approach (4–7). Therefore, we isolated and characterized salt-tolerant PGPR strains from rhizospheric soil of rice plants grown in salt-affected areas of Pindi Bhattiyan (31°53′52.12″N, 73°16′22.87″E), Hafizabad District, Punjab, Pakistan. Based on the screening process, isolate AN-PRR1 was prioritized, and its sequencing was initiated to clarify its taxonomic position, to reveal the genetic background of its characteristics, and finally to shed more light on the complete biosynthetic capacity for secondary metabolism.

Genomic DNA (gDNA) of AN-PRR1 was harvested from an overnight culture grown at 37°C in lysogeny broth (8) on a rotary shaker (220 rpm) and isolated as previously described (9). The gDNA (20 µg) was sheared using a Covaris g-TUBE device; libraries were constructed using the SMRTbell template preparation v1.0, Sequel binding v2.0, and MagBead v2 kits, followed by size selection using the BluePippin size selection system (Sage Science, Inc.). The 6-kb multiplex library was sequenced on a PacBio Sequel instrument and one SMRT cell. No quality filtering was conducted; however, subreads shorter than 50 bp were discarded. The remaining PacBio long reads were assembled using SMRT Link v7.0.1 and HGAP4 (10, 11). Default settings were used for all software, except for the HGAP4 genome size estimate parameter, which was set to 5 Mbp. Overall, the reads were assembled into one contig to construct a draft genome sequence. Since this genome resulted in a high number of pseudogenes (1,850 out of 4,592 coding DNA sequences [CDS]), mainly due to frame shifts and short indels, the gDNA extraction procedure was repeated, and upon library preparation (10 µg gDNA; Nextera XT paired-end library), the genome was resequenced using the Illumina NovaSeq 6000 platform. The initial quality assessment was based on data passing the Illumina Chastity filter. Subsequently, reads containing a PhiX control signal were removed. In addition, reads containing adapters were clipped. The second quality assessment was based on the remaining reads using the FASTQC tool v0.11.8 (12). Upon filtering, high-quality Illumina reads were assembled into contigs using ABySS v2.0.2 (13). The results of both sequencing platforms were subsequently used to perform a *de novo* hybrid assembly. The contigs were linked based on the alignment of the PacBio long reads.

Citation Nawaz A, Mubeen F, Qamar ZU, Marghoob MU, Aziz S, Gross H. 2021. Draft genome sequence of the halophilic strain *Citrobacter braakii* AN-PRR1, isolated from rhizospheric soil of rice (*Oryza sativa* L.) from Pakistan. *Microbiol Resour Announc* 10:e00787-21. <https://doi.org/10.1128/MRA.00787-21>.

Editor Leighton Pritchard, SIPBS, University of Strathclyde

Copyright © 2021 Nawaz 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 Harald Gross, harald.gross@uni-tuebingen.de.

Received 4 August 2021

Accepted 24 August 2021

Published 23 September 2021

TABLE 1 Sequencing metrics for *C. braakii* AN-PRR1

Statistic or characteristic	Data for <i>C. braakii</i> AN-PRR1
PacBio sequencing	
No. of reads	422,209
Mean read length (bp)	4,428
No. of mapped reads	417,958
Avg coverage (×)	348.95
No. of contigs	1
N_{50} (bp)	4,936,749
No. of gaps	0
Illumina sequencing	
Illumina read length (nt)	2 × 150
No. of read pairs	28,344,083
Yield (Mbp)	8,149
Avg quality	35.47
Avg coverage (×)	1,638.90
Insert size median (bp)	387.00
<i>De novo</i> hybrid assembly	
Genome size (bp)	4,929,517
GC content (%)	52.1
No. of contigs	2
N_{50} (bp)	4,102,519
No. of genes (total)	4,688
No. of genes (coding)	4,531

Alignment was performed using BLASR v1.3.1 (14). From the alignment, the orientation, order, and distance were determined using SSPACE-LongRead v1.0 (15). Using the Illumina reads, gapped regions within the contigs were closed using GapFiller v1.10 (16). Finally, assembly errors and the nucleotide disagreements between the Illumina reads and the PacBio-based sequences were corrected using Pilon v1.21 (17). The final genome sequence consists of two contigs and comprises 4.9 Mb. Genome annotation was conducted using the Prokaryotic Genome Annotation Pipeline (PGAP) v5.1 (18, 19). The genome features are summarized in Table 1.

A genome-based taxonomic analysis of strain AN-PRR1, employing the Type Strain Genome Server (TYGS) (20), revealed that *Citrobacter braakii* ATCC 51113 represents the closest related type strain. In pairwise comparisons, independent of the applied Genome BLAST Distance Phylogeny formula, the digital DNA-DNA hybridization (dDDH) values d_o , d_a , and d_g ranged from 74.5 to 91.3% and were therefore within the species threshold of 70%. Thus, AN-PRR1 represents a *Citrobacter braakii* strain. Automated secondary metabolism analysis using antiSMASH v6.0.0 (21) predicted three biosynthetic gene clusters encoding a thiopeptide, a turnerbactin-like siderophore (22), and an arylpolyene (23).

Data availability. This whole-genome sequencing (WGS) project has been deposited at DDBJ/ENA/GenBank under the accession number [JAGMWL000000000](https://accession.gtrdb.org/acc/CP030000). The corresponding raw sequencing data set has been registered in the NCBI SRA database under the accession numbers [SRX10639629](https://www.ncbi.nlm.nih.gov/sra/SRX10639629) (PacBio) and [SRX10639630](https://www.ncbi.nlm.nih.gov/sra/SRX10639630) (Illumina).

ACKNOWLEDGMENT

A.N. is grateful for a Ph.D. scholarship within the frame of the International Research Support Initiative Program (IRSIP) from the Higher Education Commission of Pakistan (HEC).

REFERENCES

1. Lesk C, Rowhani P, Ramankutty N. 2016. Influence of extreme weather disasters on global crop production. *Nature* 529:84–87. <https://doi.org/10.1038/nature16467>.
2. Hussain S, Huang J, Huang J, Ahmad S, Nanda S, Anwar S, Shakoor A, Zhu C, Zhu L, Cao X, Jin Q, Zhang J. 2020. Rice production under climate change: adaptations and mitigating strategies, p 659–686. *In* Shah F, Hasanuzzaman M, Alam M, Ullah H, Saeed M, Khan IA, Adnan M (ed), *Environment, climate, plant and vegetation growth*. Springer, Cham, Switzerland.
3. Korres NE, Norsworthy JK, Burgos NR, Oosterhuis DM. 2017. Temperature and drought impacts on rice production: an agronomic perspective regarding short- and long-term adaptation measures. *Water Resour Rural Dev* 9:12–27. <https://doi.org/10.1016/j.wrr.2016.10.001>.
4. Ullah A, Bano A, Khan N. 2021. Climate change and salinity effects on crops and chemical communication between plants and plant growth-promoting microorganisms under stress. *Front Sustain Food Syst* 5:618092. <https://doi.org/10.3389/fsufs.2021.618092>.
5. Vurukonda SSKP, Vardharajula S, Shrivastava M, SkZ A. 2016. Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. *Microbiol Res* 184:13–24. <https://doi.org/10.1016/j.micres.2015.12.003>.

6. Nawaz A, Shahbaz M, Asadullah, Imran A, Marghoob MU, Imtiaz M, Mubeen F. 2020. Potential of salt tolerant PGPR in growth and yield augmentation of wheat (*Triticum aestivum* L.) under saline conditions. *Front Microbiol* 11:2019. <https://doi.org/10.3389/fmicb.2020.02019>.
7. Tiwari S, Lata C, Chauhan PS, Nautiyal CS. 2016. *Pseudomonas putida* attunes morpho-physiological, biochemical and molecular responses in *Cicer arietinum* L. during drought stress and recovery. *Plant Physiol Biochem* 99:108–117. <https://doi.org/10.1016/j.plaphy.2015.11.001>.
8. MacWilliams MP, Liao M-K. 2006. Luria broth (LB) and Luria agar (LA) media and their uses protocol. ASM Microbe Library. American Society for Microbiology, Washington, DC.
9. Buchmann A, Gross H. 2020. Improved *de novo* draft genome sequence of the nocavionin-producing type strain *Nocardia terpenica* IFM 0706 and comparative genomics with the closely related strain *Nocardia terpenica* IFM 0406. *Microbiol Resour Announc* 9:e00689-20. <https://doi.org/10.1128/MRA.00689-20>.
10. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
11. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13:1050–1054. <https://doi.org/10.1038/nmeth.4035>.
12. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
13. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res* 27:768–777. <https://doi.org/10.1101/gr.214346.116>.
14. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238. <https://doi.org/10.1186/1471-2105-13-238>.
15. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
16. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
17. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
18. Tatusova T, DiCuccio M, Badretdin A, Chetvermin 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>.
19. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvermin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwatz 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>.
20. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
21. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>.
22. Han AW, Sandy M, Fishman B, Trindade-Silva AE, Soares CAG, Distel DL, Butler A, Haygood MG. 2013. Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont *Teredinibacter turnerae* T7901. *PLoS One* 8:e76151. <https://doi.org/10.1371/journal.pone.0076151>.
23. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158:412–421. <https://doi.org/10.1016/j.cell.2014.06.034>.