Draft Genome Sequence of *Rhodococcus rhodochrous* Strain G38GP, Isolated from the Madagascar Hissing Cockroach

Juan Guzman,[a] Andreas Vilcinskas[a,b]

aDepartment of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Giessen, Germany
bInstitute for Insect Biotechnology, Justus-Liebig-University of Giessen, Germany

ABSTRACT *Rhodococcus rhodochrous* is a bacterial species with applications in biocatalysis and bioremediation. Here, we report the draft genome sequence of strain G38GP, isolated from the gut of the cockroach *Gromphadorhina portentosa*. The genome consists of 76 contigs, with a total length of 6,256,198 bp and a GC content of 67.82%.

Fifty-two species of the genus *Rhodococcus* (*Nocardiaceae*: *Mycobacteriales*) with validly published names have been described (1). The type strain is *Rhodococcus rhodochrous* ATCC 13808[T] (=DSM 43241[T]). *R. rhodochrous* strains are used as industrial biocatalysts (2–5). Here, we report the draft genome sequence of *R. rhodochrous* strain G38GP, isolated from the gut of the Madagascar hissing cockroach, *Gromphadorhina portentosa* Schaum.

Specimens of *Gromphadorhina portentosa* are maintained as model organisms for entomology courses at the Justus-Liebig-University of Giessen, Germany. The gut of a female cockroach was dissected, suspended in sterile 50% glycerol, and cut into small fragments using sterile surgical scissors. We then plated a 1:10 dilution onto solid medium comprising 0.2% chitosan, 0.1% L-arginine, 1.5% agar, 50 mg/liter nalidixic acid, and 100 mg/liter nystatin, supplemented with trace element and vitamin solutions as previously described (6). A single colony was recultured and the 16S rRNA gene amplified and sequenced as previously reported (6). The nearly complete 16S rRNA sequence of strain G38GP (1,366 bp; GenBank accession number MW167072) showed the highest similarity to *R. rhodochrous* NBRC 16069[T] (99.85%; BBXP01000056), followed by *Rhodococcus biphenylivorans* TG9[T] (99.34%; KJ546454). The strain G38GP was deposited in the DSMZ open strain collection (DSM 111905).

A single colony was cultured in tryptic soy broth for 72 h at 28°C, and DNA was extracted from the cell pellet using the Wizard genomic DNA purification kit (Promega, Madison, WI, USA). A DNA library was prepared using the Nextera XT library preparation kit (Illumina, San Diego, CA, USA), and 150-bp paired-end reads were generated on a HiSeq 4000 device by SNPsaurus (Eugene, OR, USA). The resulting 4,344,176 read pairs were submitted to GenBank under SRA accession number SRR15212339. Default parameters were used for all software unless otherwise specified. The reads were assembled *de novo* using MaSuRCA v3.4.2 without preassembly quality control, as recommended (7). The assembly was assessed using QUAST (8). The draft genome sequence of strain G38GP was submitted to GenBank under accession number JADKNM000000000 (BioSample accession number SAMN16606023).

Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (9). The draft genome sequence of *R. rhodochrous* G38GP consists of 76 contigs, with a total length of 6,256,198 bp and an N_{50} value of 292,056 bp. The GC content is 67.82%. Genome annotation revealed a total of 5,797 genes, including 5,471 protein-coding sequences, 69 tRNA genes, and 14 rRNA genes. The genome sequence was found to be 98.9% complete using BUSCO v4.1.2 (10), based on the actinobacteria_phylum_odb10 lineage data set. Genome analysis at TYGS (11) confirmed
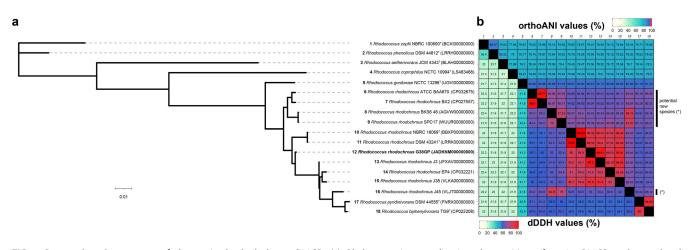
FIG 1 Genome-based taxonomy of the strain *R. rhodochrous* G38GP. (a) Phylogenomic tree showing the position of strain G38GP and 17 related *Rhodococcus* strains. (b) Overall genome relatedness metrics (orthologous ANI [orthoANI] and dDDH values) showing strain G38GP to be a member of the species *R. rhodochrous*. The phylogenomic tree was constructed using bcgTree (14) and was inferred using IQ-TREE v2.0.3 (15), with automatically optimized models for each partition and bootstraps calculated over 10⁵ generations. The dDDH and ANI values were calculated using the Genome-to-Genome Distance Calculator (12) and the OrthoANI Java tool (13).

that isolate G38GP belongs to the species *R. rhodochrous*. A phylogenomic tree showed that G38GP is closely related to the type strain *R. rhodochrous* NBRC 16069^{T} (=DSM 43241^{T} = NCTC 10210^{T}), according to the branch length calculations (Fig. 1a). Digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) values confirmed the taxonomic position of G38GP within the species *R. rhodochrous*, with values of ~85% and ~98% relative to the type strain, higher than the thresholds of 70% and 95 to 96% for species delimitation (12, 13) (Fig. 1b).

Data availability. The GenBank/EMBL/DDBJ accession numbers of the 16S rRNA gene and draft genome sequences of strain *Rhodococcus rhodochrous* G38GP (=DSM 111905) are MW167072.1 and JADKNM000000000.1, respectively. The raw sequencing data were deposited under SRA accession number SRR15212339.

REFERENCES

1. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. 2020. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70:5607–5612. https://doi.org/10.1099/ijsem.0.004332.
2. Chen B-S, Otten LG, Resch V, Muyzer G, Hanefeld U. 2013. Draft genome sequence of *Rhodococcus rhodochrous* strain ATCC 17895. Stand Genomic Sci 9:175–184. https://doi.org/10.4056/sigs.4418165.
3. Li C, Li Y, Cheng X, Feng L, Xi C, Zhang Y. 2013. Immobilization of *Rhodococcus rhodochrous* BX2 (an acetonitrile-degrading bacterium) with biofilm-forming bacteria for wastewater treatment. Bioresour Technol 131:390–396. https://doi.org/10.1016/j.biortech.2012.12.140.
4. Frederick J, Hennessy F, Horn U, de la Torre Cortés P, van den Broek M, Strych U, Willson R, Hefer CA, Daran J-MG, Sewell T, Otten LG, Brady D. 2020. The complete genome sequence of the nitrile biocatalyst *Rhodococcus rhodochrous* ATCC BAA-870. BMC Genomics 21:3. https://doi.org/10.1186/s12864-019-6405-7.
5. Komeda H, Kobayashi M, Shimizu S. 1996. Characterization of the gene cluster of high-molecular-mass nitrile hydratase (H-NHase) induced by its

reaction product in *Rhodococcus rhodochrous* J1. Proc Natl Acad Sci U S A 93:4267–4272. https://doi.org/10.1073/pnas.93.9.4267.
6. Guzman J, Sombolestani AS, Poehlein A, Daniel R, Cleenwerck I, Vandamme P, Vilcinskas A. 2021. *Entomobacter blattae* gen. nov., sp. nov., a new member of the *Acetobacteraceae* isolated from the gut of the cockroach *Gromphadorhina portentosa*. Int J Syst Evol Microbiol 71. https://doi.org/10.1099/ijsem.0.004666.
7. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677. https://doi.org/10.1093/bioinformatics/btt476.
8. 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.
9. 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 Z, 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.

10. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

11. 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.

12. Meier-Kolthoff JP, Auch AF, Klenk H-P, 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.

13. Lee I, Kim YO, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760.

14. Ankenbrand MJ, Keller A. 2016. bcgTree: automatized phylogenetic tree building from bacterial core genomes. Genome 59:783–791. https://doi.org/10.1139/gen-2015-0175.

15. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015.