Complete Genome Sequence of *Rhodococcus qingshengii* Strain VKM Ac-2784D, Isolated from *Elytrigia repens* Rhizosphere

Ivan S. Petrushin,[a,b] Yulia A. Markova,[a] Marina S. Karepova,[a] Yulia V. Zaytseva,[c] Lyudmila A. Belovezhets[d]

[a]Siberian Institute of Plant Physiology and Biochemistry, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[b]Irkutsk State University, Irkutsk, Russia
[c]Yaroslavl State University, Yaroslavl, Russia
[d]A. E. Favorsky Irkutsk Institute of Chemistry, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

ABSTRACT Rhizosphere bacteria are considered to be promising destructors of oil and its components. Bacterial species of the genus *Rhodococcus* can degrade a variety of hydrocarbons and are widely used for the bioremediation of polluted environments. Here, we report the complete genome sequence of *Rhodococcus qingshengii* strain VKM Ac-2784D.

Products of petroleum origin (oil and its components) are dangerous pollutants of the soil. Bacterial species of the genus *Rhodococcus* have the ability to degrade a variety of hydrocarbons in contaminated soil and are useful for the bioremediation of polluted environments (1–3). We isolated *Rhodococcus qingshengii* strain VKM Ac-2784D from the rhizosphere of couch grass (*Elytrigia repens*) growing at an oil-polluted site near Zalary, Irkutsk State, Russia (53°40′15″N, 102°18′29″E) (4, 5). This strain can effectively degrade oil and some model compounds (naphthalene, anthracene, and phenanthrene) (5, 6). Such activity decreases the negative effect of soil pollution and allows plants to resume growth processes (7). We suggest that the bioremediation activity of the strain in this study is associated with the production of biosurfactants and phytohormones (5).

We isolated bacteria from soil near the roots as described previously (4). To isolate bacteria, 10 g of soil near the roots was put into 100 ml sterile water and shaken for 15 min. The resulting suspension was inoculated into an accumulative medium of the following composition (g/liter): KNO_3, 4.0; KH_2PO_4, 0.6; Na_2HPO_4, 12; H_2O, 1.4; $MgSO_4$, 7; H_2O, 0.8. The medium was supplemented with only the following carbon sources: hexane, motor oil, waste motor oil, and gasoline. The strains were grown on a shaker for 30 days at 26°C. To obtain a pure culture of hydrocarbon-oxidizing bacteria, we used a solid agar medium.

DNA was extracted from the obtained bacterial strain after bead beating using the DNeasy PowerLyzer microbial kit (Qiagen, USA) according to the manufacturer's protocols. A sequence library was generated from DNA using the HyperPlus DNA sample preparation kit (KAPA Biosystems, USA). The universal bacterial primers 518F and 1064R were used to amplify the V4 to V6 hypervariable region of the bacterial 16S rRNA gene. Strain *R. qingshengii* VKM Ac-2784D showed the highest 16S rRNA gene phylogenetic affiliation to species belonging to the family *Nocardiaceae*.

Whole-genome sequencing was performed using the Illumina NextSeq platform with paired-end chemistry (2×100 bp). A total of 9,362,538 paired-end reads was obtained, giving a coverage depth of 360×. Raw read error correction and filtering was performed using the fastp tool with default settings (8). A draft assembly was built using SPAdes version 3.14.0 (9) with default settings and the "isolate" option enabled. This draft assembly contained 105 contigs with an N_{50} value of 961,557 bp and a largest contig length of 1,207,723 bp.
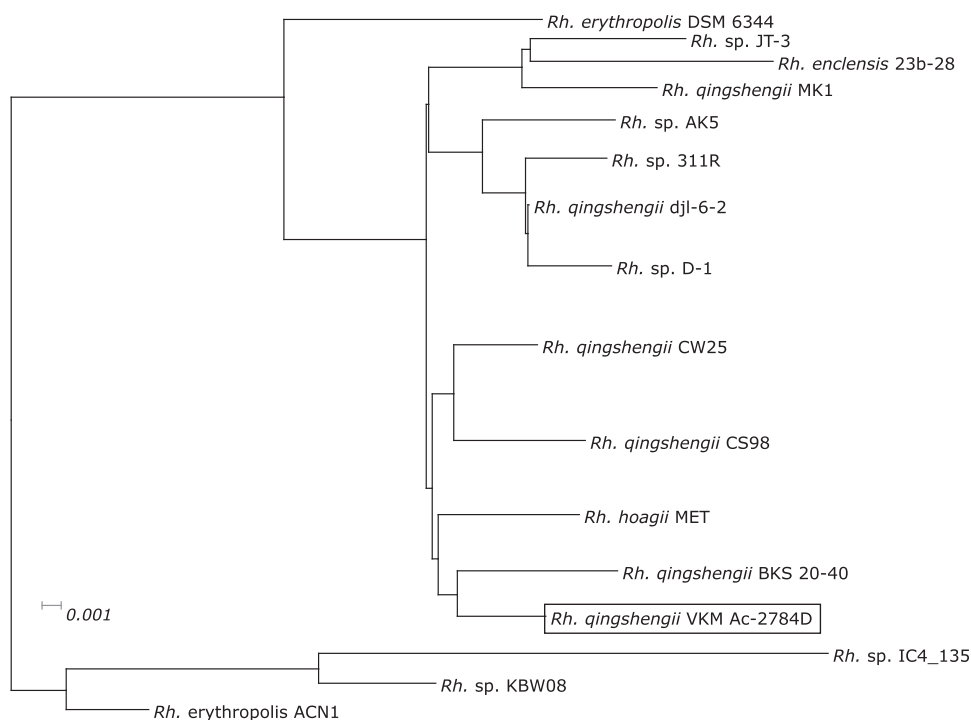
FIG 1 Phylogenetic tree of related *Rhodococcus* species.

The resulting contigs were combined into a whole chromosome using Ragout version 2.3 with default settings (10) (https://github.com/fenderglass/Ragout) using the *Rhodococcus erythropolis* SK121 whole-genome sequencing assembly (GenBank accession no. ACNO00000000) and the *Rhodococcus erythropolis* CCM2595 chromosome (NC _022115) as the reference genome sequences. The final assembly contained 1 chromosome, with a total genome sequence size of 6,251,765 bp and a GC content of 62.5%.

Genome completeness analysis with BUSCO version 5.0.0 using the data set "actinobacteria_phylum_odb10" with 356 benchmarking universal single-copy orthologs (BUSCOs) and default settings (11) gave results of 99.4% complete and single-copy, 0.6% duplicated, and no missing BUSCOs. The genome sequence was then annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (12). This assembly contained 5,775 genes encoding 5,716 protein-coding sequences, 53 tRNAs, 3 noncoding RNAs (ncRNAs), 3 rRNAs, and 49 pseudogenes, as identified by PGAP.

To compare the genomic features of *Rhodococcus qingshengii* strain VKM Ac-2784D with related species, we built a phylogenetic tree for some of these species (Fig. 1). First, we built a complete tree for >390 strains of *Rhodococcus* available in GenBank, and then we chose the 16 closest to them for a detailed tree. The tree is based on ~400 universal marker genes determined by PhyloPhlAn version 3.0 (13) (with default settings using the maximum-likelihood method).

We performed a search for biosynthetic gene clusters (BGCs) encoding secondary metabolites with the Web version of antiSMASH version 6.0 (14) and found 19 functional clusters (regions) on the chromosome, including clusters for the production of heterobactin, ectoine, erythrochelin, bacteriocin, ectoine, and heterobactin A/heterobactin S2 (nonribosomal peptide synthetase [NRPS]).

Availability of the *Rhodococcus qingshengii* strain VKM Ac-2784D genome sequence will aid in understanding its hydrocarbon-degrading mechanisms and broaden our knowledge of plant-bacteria interactions.

Data availability. This whole-genome shotgun sequence has been deposited at DDBJ/ENA/GenBank under the accession no. CP064920. The raw reads are available via

ACKNOWLEDGMENTS

REFERENCES

1. Kuyukina MS, Ivshina IB. 2010. Application of Rhodococcus in bioremediation of contaminated environments, p 231–262. *In* Alvarez H (ed), Biology of Rhodococcus. Microbiology monographs, vol 16. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-642-12937-7_9.
2. dos Santos JJ, Maranho LT. 2018. Rhizospheric microorganisms as a solution for the recovery of soils contaminated by petroleum: a review. J Environ Manage 210:104–113. https://doi.org/10.1016/j.jenvman.2018.01.015.
3. Martínková L, Uhnáková B, Pátek M, Nešvera J, Křen V. 2009. Biodegradation potential of the genus Rhodococcus. Environ Int 35:162–177. https://doi.org/10.1016/j.envint.2008.07.018.
4. Belovezhets LA, Makarova LE, Tretyakova MS, Markova YA, Dudareva LV, Semenova NV. 2017. Possible pathways for destruction of polyaromatic hydrocarbons by some oil-degrading bacteria isolated from plant endosphere and rhizosphere. Appl Biochem Microbiol 53:68–72. https://doi.org/10.1134/S0003683817010069.
5. Belovezhets LA, Tretyakova MS, Markova YA, Levchuk AA. 2020. Use of rhizosphere microorganisms for bioremediation of oil contaminated soils. IOP Conf Ser Earth Environ Sci 408:012086. https://doi.org/10.1088/1755-1315/408/1/012086.
6. Tretyakova MS, Belovezhets LA, Markova YA. 2019. Microbial preparation for bioremediation of soil contaminated with oil and oil products. Russia. https://patents.google.com/patent/RU2705290C1/en.
7. Tretyakova MS, Ivanova MV, Belovezhets LA, Markova YA. 2019. Possible use of oil-degrading microorganisms for protection of plants growing under conditions of oil pollution. IOP Conf Ser Earth Environ Sci 315:072046. https://doi.org/10.1088/1755-1315/315/7/072046.
8. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
9. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102.
10. Kolmogorov M, Armstrong J, Raney BJ, Streeter I, Dunn M, Yang F, Odom D, Flicek P, Keane TM, Thybert D, Paten B, Pham S. 2018. Chromosome assembly of large and complex genomes using multiple references. Genome Res 28:1720–1732. https://doi.org/10.1101/gr.236273.118.
11. 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.
12. 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.
13. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500. https://doi.org/10.1038/s41467-020-16366-7.
14. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.