Genome Sequence Resource of *Bacillus* sp. RRD69, a Beneficial Bacterial Endophyte Isolated from Switchgrass Plants

Zhenzhen Zhao,¹ Norbert Bokros,² Seth DeBolt,² Piao Yang,¹ and Ye Xia^{1,†}

¹ Department of Plant Pathology, The Ohio State University, Columbus, Ohio, 43210, U.S.A.

² Department of Horticulture, University of Kentucky, Lexington, Kentucky, 40546, U.S.A.

Abstract

We report here the genome sequence of *Bacillus* sp. RRD69, a plant-growth-promoting bacterial endophyte isolated from switchgrass plants grown on a reclaimed coal-mining site in Kentucky. RRD69 is predicted to contain 3,758 protein-coding genes, with a genome size of 3.715 Mbp and a 41.41% GC content.

Rhizosphere and endosphere localized bacterial communities are inextricably linked with improved crop performance. The *Bacillus* spp., a group of plant-growth-promoting bacteria (PGPB), are considered among the most well-studied, effective, and commercially important PGPB, with multiple members developed as biofertilizers and biocontrol agents in agricultural production (Hashem et al. 2019; Radhakrishnan et al. 2017). Switchgrass (*Panicum virgatum* L.), a widely planted forage crop and ground cover plant, attracts more attention as an important biofuel crop (Xia et al. 2013). We report here the genome information of the bacterial endophyte *Bacillus* sp. RRD69 associated with switchgrass plants, which was reported to display beneficial potential for increasing the growth and weight of switchgrass plants (Xia et al. 2013).

In our study, the switchgrass samples were collected from a reclaimed coal-mining site in Kentucky (Xia et al. 2013). The roots of the collected plant samples were cut into 3- to 5-cm-long segments, surface sterilized with the 20% bleach solution for 15 min, and rinsed with sterilized water five times. The surface-sterilized segments were further cut into smaller segments of 1 to 1.5 cm and placed on plates with tryptic soy agar medium (Sigma-Aldrich). These plates were incubated in a 26°C incubator for 3 to 5 days. Individual isolates emerging from those segments were separately isolated and cultured for further DNA extraction. The 16S ribosomal DNA amplifications with primers 27f (5'-GAGTTTGATCCTGGCTCA-3') and 1498r (5'-ACGGCTACCTTGTACGACTT-3') were carried out. The amplified PCR products were further sequenced and analyzed by BLASTn searches in the NCBI databases (Devulder et al. 2003; Mignard and Flandrois 2006). The top hits as the most possible taxonomic species were identified. Among all the isolates, the isolate of *Bacillus* sp. RRD69 was identified (Xia et al. 2013). This isolate was further cultured in tryptic soy broth media on a rotary shaker

Funding

The project was supported under the Department of Energy contract DE-AC02-05CH11231. This project was also partially supported by the Hatch Project from the United States Department of Agriculture–National Institute of Food and Agriculture (USDA-NIFA-OHO01392) at The Ohio State University and OARDC SEED GRANT OHOA1615.

Keywords

Bacillus, bacteria-plant symbiosis, biocontrol, endophytes, genome information, sequencing, switchgrass

†Corresponding author: Y. Xia; xia.374@osu.edu

*The e-Xtra logo stands for “electronic extra” and indicates that supplementary materials are published online.

The author(s) declare no conflict of interest.

Accepted for publication 23 July 2021.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

Fig. 1. Circular representation of the *Bacillus* sp. RRD69 genome using Circos. Circles, from outside to inside, denote **A**, protein-coding genes colored by size; **B**, RNA genes; **C**, transmembrane helix regions; **D**, GC content along with a 1-kb window, with red lines indicating the regions above 41.4% genome average and black lines indicating the regions below the genome average; **E**, GC skew, with red lines indicating a skew greater than zero and black lines indicating a skew less than zero; and **F**, genes annotated into distinct PANTHER protein classes. The repository for the storage of scripts used to construct the figure can be found online.

overnight at 26°C. The bacterial cells were subject to the cetyltrimethylammonium bromide genomic DNA extraction (Wilson 2001) and genome sequencing (Xia et al. 2013).

A PacBio SMRTbell library was constructed and sequenced with the PacBio RS platform. In total, 512,579 raw reads were initially generated. After quality filtering, trimming, and assembling using the Hierarchical Genome Assembly Process (v.2.3.0) with the default settings (Chin et al. 2013), the final draft assembly contained five contigs in five scaffolds, with a total size of 3.715 Mbp and an N_{50} value of 975.148 kb. The input read coverage was 191.9x, and the total GC content was 41.41%. Prodigal (v.2.5) and GenePRIMP pipeline were used for the subsequent genome annotation (Hyatt et al. 2010; Pati et al. 2010). Genome annotation predicted 3,884 genes, which included 3,758 predicted protein-coding genes. These protein-coding genes were used to search through the NCBI nonredundant database, UniProt, TIGRFam, Pfam, Kyoto Encyclopedia of Genes and Genomes, Clusters of Orthologous Genes, PANTHER, and InterPro databases (Finn et al. 2014; Haft et al. 2003; Kanehisa and Goto 2000; Pruitt et al. 2007; Tatusov et al. 2003; The UniProt Consortium 2019; Thomas et al. 2003). For the remaining 126 genes, 26 ribosomal RNA (rRNA) genes, 73 transfer RNA (tRNA) genes, and 27 noncoding RNAs were identified by using rRNA genes models from SILVA (Pruesse et al. 2007), the tRNAscan-SE tool (v.2.0.5), and Rfam profiles via INFERNAL (v.1.1.3), respectively (Finn et al. 2014). CheckM (v.1.0.18) was used to estimate the completeness of the genome at 99.6% with a predicted contamination level of only 0.6% (Arkin et al. 2018; Parks et al. 2015). Using the PANTHER hidden Markov model (HMM) scoring tool panther-Score (v.2.2), the protein sequences were further mapped against the PANTHER HMM database (v.15) to functionally annotate the genes and query for the significantly overrepresented genes (Mi et al. 2019). A circular representation of

selected annotations and genome characteristics generated using the Circos software is depicted in Figure 1 (Krzywinski et al. 2009). Additional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes (IMG) platform developed by the Joint Genome Institute (JGI, Walnut Creek, CA, U.S.A.) (Markowitz et al. 2012).

A previous study has reported that multiple endophytic bacteria associated with switchgrass, including several *Bacillus* strains, displayed plant-growth-promoting activity (Xia et al. 2013). In addition, the genome sequencing of one *Bacillus* strain, *Bacillus* sp. YF23, was reported (Xia et al. 2019). The goal of this study was to address the genome information of another beneficial endophyte, *Bacillus* sp. RRD69. The comparative analysis of *Bacillus* sp. RRD69 with *Bacillus* sp. YF23 is provided in Supplementary Table S1. With the emerging *Bacillus* PGPB genomic resources, we will have an improved understanding of *Bacillus* PGPB and their associations with plants.

Data Availability

The whole-genome sequence of this bacterium has been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the BioProject accession number PRJNA322990. The associated NCBI BioSample and NCBI Sequence Read Archive accession numbers are SAMN05216491 and SRP088163. The associated sequence data, as well as further information on sample preparation, genome assembly, and annotation, can be found at JGI with IMG taxon ID number 2681812863 and NCBI ID 1855345. The repository for the storage of scripts used to construct Figure 1 can be found online. The version described in this article is the first version.

Acknowledgments

The genome sequencing and data annotation were carried out in the United States Department of Energy (DOE) JGI, a DOE Office of Science User Facility.

Author-Recommended Internet Resource

Repository for the storage of scripts: <https://github.com/nbo245/rrd69>

Literature Cited

- Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., Glass, E. M., Chivian, D., Gunter, D., Weston, D. J., Allen, B. H., Baumohl, J., Best, A. A., Bowen, B., Brenner, S. E., Bun, C. C., Chandonia, J. M., Chia, J. M., Colasanti, R., Conrad, N., Davis, J. J., Davison, B. H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J. N., Fang, G., Faria, J. P., Frybarger, P. M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H. L., He, F., Jain, R., Joachimiak, M. P., Keegan, K. P., Kondo, S., Kumar, V., Land, M. L., Meyer, F., Mills, M., Novichkov, P. S., Oh, T., Olsen, G. J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S. S., Price, G. A., Ramakrishnan, S., Ranjan, P., Ronald, P. C., Schatz, M. C., Seaver, S. M. D., Shukla, M., Sutormin, R. A., Syed, M. H., Thomason, J., Tintile, N. L., Wang, D., Xia, F., Yoo, H., Yoo, S., and Yu, D. 2018. KBase: The United States Department of Energy systems biology knowledgebase. *Nat. Biotechnol.* 36:566-569.
- Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10:563-569.
- Devulder, G., Perrière, G., Baty, F., and Flandrois, J. P. 2003. BIBI, a bioinformatics bacterial identification tool. *J. Clin. Microbiol.* 41:1785-1787.
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. 2014. Pfam: The protein families database. *Nucleic Acids Res.* 42:D222-D230.
- Haft, D. H., Selengut, J. D., and White, O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31:371-373.
- Hashem, A., Tabassum, B., and Abd-Allah, E. F. 2019. *Bacillus subtilis*: A plant-growth promoting rhizobacterium that also impacts biotic stress. *Saudi J. Biol. Sci.* 26:1291-1297.
- Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11:119.
- Kanehisa, M., and Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27-30.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res.* 19:1639-1645.
- Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N., and Kyrpides, N. C. 2012. IMG: The Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* 40:D115-D122.
- Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P. D. 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* 14: 703-721.
- Mignard, S., and Flandrois, J. P. 2006. 16S rRNA sequencing in routine bacterial identification: A 30-month experiment. *J. Microbiol. Methods* 67:574-581.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25: 1043-1055.

- Pati, A., Ivanova, N. N., Mikhailova, N., Ovchinnikova, G., Hooper, S. D., Lykidis, A., and Kyrpides, N. C. 2010. GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* 7:455-457.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., and Glöckner, F. O. 2007. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188-7196.
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35:D61-D65.
- Radhakrishnan, R., Hashem, A., and Abd-Allah, E. F. 2017. *Bacillus*: A biological tool for crop improvement through bio-molecular changes in adverse environments. *Front. Physiol.* 8:667.
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. 2003. The COG database: An updated version includes eukaryotes. *BMC Bioinf.* 4:41.
- The UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47:D506-D515.
- Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. 2003. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* 13:2129-2141.
- Wilson, K. 2001. Preparation of genomic DNA from bacteria. *Curr. Protoc. Mol. Biol.* 56:2.4.
- Xia, Y., DeBolt, S., Ma, Q., McDermaid, A., Wang, C., Shapiro, N., Woyke, T., and Kyrpides, N. C. 2019. Improved draft genome sequence of *Bacillus* sp. strain YF23, which has plant growth-promoting activity. *Microbiol. Resour. Announce.* 8:00099-19.
- Xia, Y., Greisworth, E., Mucci, C., Williams, M. A., and De Bolt, S. 2013. Characterization of culturable bacterial endophytes of switchgrass (*Panicum virgatum* L.) and their capacity to influence plant growth. *GCB Bioenergy* 5:674-682.