Draft Whole-Genome Sequence of *Sphingobium* sp. Strain PNB, a Versatile Polycyclic Aromatic Hydrocarbon-Degrading Bacterium

Madhumita Roy,ª Tapan K. Duttaª

ªDepartment of Microbiology, Bose Institute, Kolkata, India

ABSTRACT *Sphingobium* sp. strain PNB can completely degrade phenanthrene, naphthalene, and biphenyl as the sole carbon and energy source. The strain is also capable of cometabolizing benzo[a]pyrene, pyrene, acenaphthene, fluoranthene, etc. Here, we report the 5.69-Mb assembly and annotation of the genome sequence of strain PNB, obtained using Illumina sequencing.

*S*phingobium sp. strain PNB is a Gram-negative, rod-shaped, chemoheterotrophic, strictly aerobic bacterium that forms yellow colonies. It was isolated from a municipal waste-contaminated soil sample from Dhapa, Kolkata, India (22.5373°N, 88.4334°E), and selected due to its versatile capability to completely degrade various aromatics, like phenanthrene, naphthalene, and biphenyl, as the sole source of carbon and energy (1). Strain PNB can also partially utilize anthracene as the sole carbon source, producing 3-hydroxy-2-naphthoic acid as the dead-end metabolite, and is able to cometabolize a number of high-molecular-weight polycyclic aromatic hydrocarbons (PAHs) in the presence of phenanthrene (2–4). Thus, due to its versatile catabolic potential regarding various PAHs and other aromatics, including central metabolites, information about the genome sequence of the strain would be helpful in future for directed strain improvement by targeted genetic engineering for biotechnological applications, including PAH bioremediation.

For molecular characterization, *Sphingobium* sp. strain PNB, originally isolated from soil and maintained in the lab, was grown in lysogeny broth (LB) medium (5) at 29°C for 24 h, and high-quality DNA was extracted using the GeneJET genomic DNA purification kit (Thermo Fisher Scientific, MA, USA). The concentration and quality of the extracted DNA were determined using the PHERAstar FSX multimode microplate reader (BMG Labtech). The genomic DNA was subjected to sequencing via Illumina (MiSeq platform) with paired-end technology (2×250 bp), and barcode strategies were applied according to the manufacturer's instructions (Illumina, USA). A sequencing library was prepared using the Nextera XT DNA library preparation kit (Illumina). Both paired-end sequencing read files were separately subjected to the FastQC v0.115 tool, and the necessary trimming was executed using the FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). The most suitable kmer size was estimated in the processed reads of the *Sphingobium* genome by applying KmerGenie v1.7051 software (6). In this analysis, the best kmer size obtained was 85. The SPAdes v3.15.0 assembler was utilized for *de novo* assembly using the good-quality processed reads (7). This analysis was performed using the "careful" option, which is recommended for mismatch correction. A total of 332 scaffolds were constructed in the *de novo* assembly. After the removal of some small scaffolds (<500 bp), a total of 242 scaffolds were obtained. The exact N_{50} value was 49,683 bp, and the sequencing coverage was $10\times$. In all, 5,696,236 bp (~5.7 Mb) of assembled consensus sequences were obtained, with a GC content of 64.11%. Functional annotation of the genome was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html) (8). The total number of genes and coding DNA sequences (CDS) annotated were 5,553

Address correspondence to Tapan K. Dutta, tapan@jcbose.ac.in.

and 5,491, respectively. In the above analyses, default parameters were used for all software unless otherwise specified.

The genome harbors pathways for the degradation of PAHs/biphenyl, catechol (*ortho-* and *meta*-cleavage), homogentisate, salicylate, phenol (containing multicomponent hydroxylase), nonoxidative decarboxylase-mediated aromatic acid, etc. In addition, chromium, copper, and arsenic resistance genes were found in the genome, suggesting the suitability of the strain for the bioremediation of diverse PAH-contaminated sites.

Data availability. The raw reads are available under NCBI BioProject accession number PRJNA757604. The raw reads are available under SRA accession number SRX11971035. The *Sphingobium* sp. strain PNB whole-genome shotgun sequencing project has been deposited in DDBJ/ENA/GenBank under accession number JAINEZ000000000. The version described in this paper is JAINEZ010000000.

REFERENCES

1. Roy M, Khara P, Dutta TK. 2012. *meta*-Cleavage of hydroxynaphthoic acids in the degradation of phenanthrene by *Sphingobium* sp. strain PNB. Microbiology (Reading) 158:685–882. https://doi.org/10.1099/mic.0.053363-0.
2. Roy M, Khara P, Basu S, Dutta TK. 2013. Catabolic versatility of *Sphingobium* sp. strain PNB capable of degrading structurally diverse aromatic compounds. J Bioremediat Biodegrad 4:173. https://doi.org/10.4172/2155-6199.1000173.
3. Khara P, Roy M, Chakraborty J, Ghosal D, Dutta TK. 2014. Functional characterization of diverse ring-hydroxylating oxygenases and induction of complex catabolic gene clusters in multiple aromatics degrading *Sphingobium* sp. strain PNB. FEBS Open Bio 4:290–300. https://doi.org/10.1016/j.fob.2014.03.001.
4. Khara P, Roy M, Chakraborty J, Dutta A, Dutta TK. 2018. Characterization of a topologically unique oxygenase from *Sphingobium* sp. PNB capable of catalyzing a broad spectrum of aromatics. Enzyme Microb Technol 111:74–80. https://doi.org/10.1016/j.enzmictec.2017.10.006.
5. Luria SE, Williams RC, Backus RC. 1951. Electron micrographic counts of bacteriophage particles. J Bacteriol 61:179–188. https://doi.org/10.1128/jb.61.2.179-188.1951.
6. Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30:31–37. https://doi.org/10.1093/bioinformatics/btt310.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
8. Pruitt KD, Tatusova T, Klimke W, Maglott DR. 2009. NCBI reference sequences: current status, policy and new initiatives. Nucleic Acids Res 37:D32–D36. https://doi.org/10.1093/nar/gkn721.