

Short Communication

Unveiling the draft genome sequence of diesel-degrading *Paenibacillus* sp. strain d9, a surfactant producer isolated from diesel-contaminated soil

Vikas Sharma^{1,2*}, Roshini Govinden¹ and Johnson Lin¹

¹Department of Microbiology, School of Life Sciences, University of KwaZulu-Natal (Westville), Private Bag X 54001, Durban-4000, Republic of South Africa

²Department of Biotechnology, Ambala College of Engineering and Applied Research, Ambala Cantt, Jagadhari Rd, P.O, Sambhalkha, Haryana 133101, India

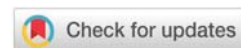
Received: 18 July, 2023
Accepted: 08 August, 2023
Published: 09 August, 2023

*Corresponding author: Dr. Vikas Sharma, Department of Biotechnology, Ambala College of Engineering and Applied Research, Ambala Cantt, Jagadhari Rd, P.O, Sambhalkha, Haryana 133101, India, Tel: +918708633613; E-mail: sindhurvikas@gmail.com

Keywords: *Paenibacillus*; Diesel degradation; Surfactant; Draft genome

Copyright License: © 2023 Sharma V, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://www.peertechzpublications.org>



Abstract

Introduction: Gram-positive bacteria, particularly *Bacillus* and *Paenibacillus* spp., have gained significant attention for their potential in environmental bioremediation (biosurfactant production) and diverse biotechnological applications. Among these, *Paenibacillus* sp. D9, isolated from oil-contaminated soil, has shown diesel and engine oil degradation capabilities and biosurfactant production. However, its role in alkane degradation remains unexplored.

Methodology and Result: To shed light on its unique attributes, we conducted whole-genome sequencing of *Paenibacillus* sp. D9 using the Illumina HiSeq 2000 platform. The draft genome comprised 56 contigs and 7 scaffolds, with a size of 5,645,302 bp at 157.94× coverage and a G + C content of 58.13%. A total of 9,950 Coding Sequences (CDSs) were predicted, and functional annotation revealed 3,283 (43.19%) and 3,155 (58.8%) putative genes based on Bacterial Annotation System (BASys) and Rapid Annotation using the Subsystems Technology (RAST) subsystem categorization, respectively. Furthermore, 93 tRNA and 23 rRNA genes were identified.

Conclusion: This genome announcement provides valuable insights into the genetic potential of *Paenibacillus* sp. D9 and paves the way for future research in its biotechnological applications.

Genome announcement

Gram-positive bacteria particularly *Bacillus* and *Paenibacillus* spp. have been attracting interest in both environmental bioremediation strategies (biosurfactant production) and biotechnological applications [1,2]. *Paenibacillus* spp. are known to be associated with a wide variety of applications ranging from agriculture and horticulture to industrial and medicine [3]. *Paenibacillus* sp. D9 [4] isolated from oil-contaminated soil has been reported to degrade diesel and produce biosurfactants; however, its role in alkane degradation has not been established. The uniqueness of this organism lies in the presence of commercially significant proteins and biosurfactant-producing genes, in addition to its diesel

and engine oil degradation capabilities. Therefore, genome sequencing was conducted. The whole-genome sequencing was performed using the Illumina HiSeq 2000 platform by the Beijing Genomics Institute (BGI), Shenzhen, China. Paired-end libraries with average insert sizes of ~500 base pair (bp) and ~6 kilobases (kb) were generated following the manufacturer's instructions. The reads were aligned with the reference sequence using SOAPaligner (version 2.21) software to assess differences between the sequencing and reference species in terms of average depth and coverage ratio [5]. The initial processing involved de novo assembly of the filtered short reads using SOAPdenovo v 2.04, following the method described previously [6]. Subsequently, contigs were manually connected based on their 500 bp and 6 kb paired-end relationships. The



resulting draft genome comprised 56 contigs and 7 scaffolds, with a maximum contig size of 582,249 bp and a scaffold size of 5,594,870 (Table 1). The genome size was 5,645,302 bp at $157.94 \times$ coverage, with N_{50} of 5,594,870 bp and N_{90} of 5,594,870 bp and the G + C content was 58.13%. A total of 9,950 coding DNA sequences (CDSs) or Open Reading Frames (ORFs) were predicted using Glimmer v3.02 [7] and homologous comparison to a non-redundant public database was performed by Basic Local Alignment Search Tool (BLAST) for function annotation. The genome annotation was performed using the BASys server and the output was downloaded in GenBank format resulting in 7,600 (CDSs) [8]. The genome was further annotated with RAST server [9]. The features for Draft genome sequence of *Paenibacillus* sp. D9 are summarized in Table 1. Among the 7,600 predicted protein-coding genes, 3,283 (43.19%) have putative functions based on BASys subsystem categorization, while the RAST server assigned functions to 3,155 genes (58.8%) among 5,359 predicted. Additionally, 93 tRNA genes covering all 20 amino acids were identified using the tRNAscan-SE program [10] and 23 rRNA genes were identified using RNAmmer [11]. The contigs were submitted to GeneBank, and NCBI published the sequence data in April 2015.

Nucleotide sequence accession number

The nucleotide sequence accession numbers for this Whole Genome Sequencing (WGS) project are as follows: The project has been deposited at DDBJ/EMBL/GenBank under the accession JZJ00000000. The version described in this paper is JZJ01000000. The Bioproject is registered under accession: PRJNA277007 ID: 277007. The *Paenibacillus* sp. D9 isolate has been deposited at the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures and is available under the Accession No. DSM 101888.

Table 1: General features of the *Paenibacillus* sp. strain D9 genome.

Sl. No	Features	Values
1	Genome Size (bp)	5,594,870
2	DNA G+C content (%)	58.13
3	Total contigs	56
4	Total genes (CDS)	9,950
5	rRNA	23
6	tRNA	93
7	Genes with predicted function	7,600

References

- Larkin MJ, Kulakov LA, Allen CC. Biodegradation and Rhodococcus--masters of catabolic versatility. *Curr Opin Biotechnol.* 2005 Jun;16(3):282-90. doi: 10.1016/j.copbio.2005.04.007. PMID: 15961029.
- Najafi AR, Rahimpour MR, Jahanmiri AH, Roostaazad R, Arabian D, Soleimani M, Jamshidnejad Z. Interactive optimization of biosurfactant production by *Paenibacillus alvei* ARN63 isolated from an Iranian oil well. *Colloids Surf B Biointerfaces.* 2011 Jan 1;82(1):33-9. doi: 10.1016/j.colsurfb.2010.08.010. Epub 2010 Aug 13. PMID: 20846835.
- Lal S, Tabacchioni S. Ecology and biotechnological potential of *Paenibacillus*

polymyxa: a minireview. *Indian J Microbiol.* 2009 Mar;49(1):2-10. doi: 10.1007/s12088-009-0008-y. Epub 2009 Apr 21. PMID: 23100748; PMCID: PMC3450047.

- Ganesh A, Lin J. Diesel degradation and biosurfactant production by Gram-positive isolates. *Afr J Biotechnol.* 2009;8(21):5847-54. doi: 10.5897/AJB09.811.
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics.* 2009 Aug 1;25(15):1966-7. doi: 10.1093/bioinformatics/btp336. Epub 2009 Jun 3. PMID: 19497933.
- Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics.* 2008 Mar 1;24(5):713-4. doi: 10.1093/bioinformatics/btn025. Epub 2008 Jan 28. PMID: 18227114.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics.* 2007 Mar 15;23(6):673-9. doi: 10.1093/bioinformatics/btm009. Epub 2007 Jan 19. PMID: 17237039; PMCID: PMC2387122.
- Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W455-9. doi: 10.1093/nar/gki593. PMID: 15980511; PMCID: PMC1160269.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008 Feb 8;9:75. doi: 10.1186/1471-2164-9-75. PMID: 18261238; PMCID: PMC2265698.
- Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W686-9. doi: 10.1093/nar/gki366. PMID: 15980563; PMCID: PMC1160127.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35(9):3100-8. doi: 10.1093/nar/gkm160. Epub 2007 Apr 22. PMID: 17452365; PMCID: PMC1888812.

Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROME0, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services (<https://www.peertechz.com/submission>).

Peertechz journals wishes everlasting success in your every endeavours.