CrossMark
click for updates

Draft Genome Sequence of *Acinetobacter baumannii* IITR88, a Bacterium Degrading Indoles and Other Aromatic Compounds

Raj Kumar Regar,[a] Vivek Kumar Gaur,[a] Gayatri Mishra,[b] Sudhir Jadhao,[b] Mohan Kamthan,[a] Natesan Manickam[a]

Environmental Biotechnology Laboratory, Environmental Toxicology Group, CSIR Indian Institute of Toxicology Research (IITR), Vishvigyan Bhawan, Lucknow, India[a];
Bionivid Technology Pvt. Ltd., Bengaluru, Karnataka, India[b]

This article represents CSIR-IITR communication no. 3374.

Here, we report the 4.16-Mb draft genome sequence of an indole-degrading bacterium, *Acinetobacter baumannii* IITR88, isolated from the Bhagirathi river in India. A total of 4,069 coding regions (CDSs), 3 rRNAs, and 52 tRNAs were predicted. Genes for the degradation of indoles, phenylacetaldehyde, anthranilate, and several other aromatic compounds were present.

Address correspondence to Natesan Manickam, nmanickam@iitr.res.in.

The genome of *Acinetobacter baumannii* IITR88, a Gram-negative and aerobic bacterium, was sequenced here. Illumina platform paired-end technology produced a total of 33,357,632 paired-end reads of 151 bp. We used the NGS QC toolkit version 2.3 (1) to filter high-quality (HQ) data (cutoff read length for HQ, 70%; cutoff quality score, 20) for genome assembly. A total of 29,571,406 reads were used for the assembly with Velvet 1.2.08 (2) (at a k-mer length of 69). Based on the paired-end directional information, the genome was assembled into 43 scaffolds, with an N_{50} of 270,137 bp and average scaffold length of 96,871 bp using the SSPACE version 3.0 (3) scaffolder. The genome was assembled, resulting in a genome size of ~4.16 Mb, with a G+C content of 38.97%.

The input sequences assembled into 43 scaffolds and were anchored onto the nearest reference genome of *A. baumannii* AB5256 using CONTIGuator (4). The final genome draft of strain IITR88 consists of a pseudochromosome and unaligned scaffolds of length ≥218 bp. The final genome draft of strain IITR88 consists of 22 scaffolds, with an average length of 181,193.87 bp and N_{50} size of 3,990,690 bp, constituting a genome size of 4,167,459 bp.

The draft genome (22 scaffolds), comprising 4,167,459 nucleotides, was annotated with the help of the Rapid Annotations using Subsystems Technology (RAST) version 2.0 (5), Prodigal version 2.6.2 (6), and RNAmmer 1.2 servers (7). Three types of rRNAs and possible tRNAs were identified, indicating a high degree of completeness in the genome assembly. A total of 4,077 genes predicted from Prodigal, 4,069 predicted coding regions (CDSs), and 3 rRNAs (16S rRNA, 28S rRNA, and 5S rRNA) were predicted. ARAGORN version 1.2.36 (8) predicted 52 tRNAs. No plasmid was identified when it was analyzed using Webcutter (version 2.0) and PlasmidFinder (version 1.3) (9), and 3 prophages were identified using the PHAST server (10). Taxonomic identification using EzTaxon (11) and MEGA6 revealed that *Acinetobacter venetianus* and *Acinetobacter junii* are putative

species, and *A. baumannii* is the closest homolog of the assembled genome.

RAST annotation shows that *A. baumannii* strains AB5256 (score, 539), AB4857 (score, 516), and TYTH-1 (score, 500) are the closest neighbors of *A. baumannii* IITR88. RAST annotation also shows that strain IITR88 contains the 3,072 characterized proteins, 1,046 hypothetical/putative proteins, and 1,146 proteins with Pathway Annotation. Strain IITR88 has genes coding for the enzymes indole-3-glycerol phosphate synthase (EC 4.1.1.48), catechol 1,2-dioxygenase (EC 1.13.11.1), benzoate 1,2-dioxygenase alpha-subunit (EC 1.14.12.10), phenylacetaldehyde dehydrogenase (EC 1.2.1.39), protocatechuate 3,4-dioxygenase (EC 1.13.11.3), butyryl-coenzyme A dehydrogenase (EC 1.3.99.2), aldehyde dehydrogenase (EC 1.2.1.3), and acetyl-coenzyme A (CoA) acetyltransferase (EC 2.3.1.9). Also, the genes for antibiotic resistance to spectinomycin 9-*O*-adenylyltransferase, β-lactamase (EC 3.5.2.6), multidrug resistance transporter, Bcr/CflA family and resistance-nodulation-division (RND) efflux systems, outer membrane lipoprotein, arsenate reductase (EC 1.20.4.1), and cobalt-zinc-cadmium resistance protein (EC 6.3.2.12) were found.

Nucleotide sequence accession numbers. The draft genome sequence of *A. baumannii* IITR88 is available in DDBJ/EMBL/GenBank under the accession no. LQAR00000000. The version described in this paper is the first version, LQAR01000000.

REFERENCES

1. **Patel RK, Jain M.** 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7:e30619. http://dx.doi.org/10.1371/journal.pone.0030619.
2. **Daniel R, Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo*

short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

3. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W**. 2011. Scaffolding pre-assembled contigs using SSPACE. Oxford J Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.

4. **Galardini M, Biondi EG, Bazzicalupo M, Mengoni A**. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med **6:**11. http://dx.doi.org/10.1186/1751-0473-6-11.

5. **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**. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.

6. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.

7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW**. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.

8. **Laslett D, Canback B**. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res **32:**11–16. http://dx.doi.org/10.1093/nar/gkh152.

9. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Aarestrup FM, Hasman H**2014. PlasmidFinder and pMLST: *in silico* detection and typing of plasmids. Antimicrob Agents Chemother **58:**3895–3903. http://dx.doi.org/10.1128/AAC.02412-14.

10. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS**. 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39:**W347–W352. http://dx.doi.org/10.1093/nar/gkr485.

11. **Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won S, Chun J**. 2012. Introducing EzTaxon: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol **62:**716–721. http://dx.doi.org/10.1099/ijs.0.038075-0.