



Draft Genome Sequence of *Idiomarina* sp. Strain 5.13, a Highly Stress-Resistant Bacterium Isolated from the Southwest Indian Ridge

Rupesh Kumar Sinha, K. P. Krishnan, P. John Kurian

National Centre for Antarctic and Ocean Research, Head land Sada, Vasco-da-Gama, Goa, India

ABSTRACT *Idiomarina* sp. strain 5.13, able to produce biopolymer and exopolysaccharide, was isolated from a sediment sample collected from the Southwest Indian Ridge, Indian Ocean. Analysis of its draft genome sequence provides insights into its remarkable stress tolerance and offers the genetic basis for harnessing the biotechnological potential of this strain.

Idiomarina sp. strain 5.13 was isolated from a sediment sample collected from the Southwest Indian Ridge (26°56.49'S, 67°19.477'E), Indian Ocean, and was found to produce biopolymer and exopolysaccharide (R. K. Sinha, K. P. Krishnan, and P. J. Kurian, unpublished data). Here, we report the draft genome sequence and annotation of strain 5.13. The genomic sequence was determined using the Illumina reversible terminator technology. The total sequence of 1,757 Mb with 6,955,430 mate-pair and 9,121,892 paired-end reads was generated from the HiSeq 2500 platform. During preprocessing, adapter sequences were removed using Cutadapt version 1.9 (1). All low-quality ($Q < 20$) data were filtered out using Sickle version 1.33. Contigs lesser than 200 bp and duplicate reads were discarded using Fastuniq (2). KmerGenie (3) was used to predict the optimal k value and assembly size, which was found to be 31 and 7,352,014 bp, respectively. *De novo* assembly was performed with Velvet version 1.2.10 (4) using default options, except the k value, which resulted in 87 contigs.

Idiomarina sp. strain 5.13 has genome length of ~4.5 Mb with a G+C content of 46.1% and no plasmids, as deduced with PlasmidFinder (5). The assembly was annotated using Prokka (6), which predicts coding DNA sequence (CDS) using Prodigal (7). A total of 4,043 CDSs were predicted. The predicted CDSs were annotated using the UniProt database (<http://www.uniprot.org/>) of the BLAST program, followed by comparison with the HMM profile database and Pfam database (8) using the HMMER3 program (9). Matches with an E value of $\leq 10^{-6}$ were retained for further annotation. A total of 1,049 hits were labeled as hypothetical proteins by the Prokka program. tRNA and transfer-messenger RNA (tmRNA) genes predicted from the contigs using Prokka, which uses Aragorn (10), found 52 tRNA genes and one tmRNA gene. The genome sequence of strain 5.13 showed the presence of phage remnant-like transcription regulator AlpA and phage packaging genetic machinery (*gpAZ nu1*). The presence of 66 transposases and phage integrase family protein scattered throughout the genome supported the likelihood of active genetic rearrangement of genome of strain 5.13. The bacterium is equipped with resistance genes against heavy metals, such as mercury (*merACPR*), cobalt, zinc, and cadmium (*czcAB*, *zur1*, *zur2*, *zntB*, *zupT*, *zitB*, and *zip* family), arsenic (*arsC*), copper (*copABCD*), nickel (*nikA*), and manganese (*mntHP*), and possesses a cation efflux genetic system (*cusAB* and *cutA*). It contained 69 genes for chemotaxis and flagellar motility. *Idiomarina* sp. strain 5.13 can accumulate polyhydroxybutyrate in intracellular storage granules (R. K. Sinha, K. P. Krishnan, and P. J. Kurian, unpublished

Received 5 January 2017 Accepted 11 January 2017 Published 9 March 2017

Citation Sinha RK, Krishnan KP, Kurian PJ. 2017. Draft genome sequence of *Idiomarina* sp. strain 5.13, a highly stress-resistant bacterium isolated from the Southwest Indian Ridge. Genome Announc 5:e01747-16. <https://doi.org/10.1128/genomeA.01747-16>.

Copyright © 2017 Sinha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Rupesh Kumar Sinha, kineto.magnetic@gmail.com, or K. P. Krishnan, kpkrishnan@gmail.com.

This article is Esso-NCAOR contribution number 02/2017.

data) conferred by the presence of *phbC* and a phasin protein family. Genetic components for exopolysaccharide production and biofilm formation and dispersion (*exoD*, *ycgF1*, *ycgF2*, *bigR*, and *bdlA123*) were also present. Genes that could allow strain 5.13 to cope with oxidative stress and cold shock (*phoH*, stringent starvation protein family, *cspAGD*) were observed, along with eight chaperone-encoding genes.

The overall genome analysis revealed a high potential of strain 5.13 in bioremediation and biotechnological application and provides insight into its adaptability to extreme environmental conditions.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [MSEI00000000](#). The version described in this paper is version MSEI01000000.

ACKNOWLEDGMENTS

We thank the Director, Esso-NCAOR, for his interest in this work. This work was undertaken as part of the project “Study of seafloor hydrothermal system in mid oceanic ridges.”

We declare no conflict of interests.

REFERENCES

1. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
2. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. 2012. FastUniq: a fast *de novo* duplicates removal tool for paired short reads. *PLoS One* 7:e52249. <https://doi.org/10.1371/journal.pone.0052249>.
3. 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>.
4. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
5. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
6. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
7. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
8. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
9. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
10. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.