

Draft genomes of five putative biosurfactant-producing *Serratia* sp. isolates from Laguna de Bay, Philippines

Nacita B. Lantican,¹ Glaezel Angelique T. Barredo,¹ Albert R. Rosana,¹ Ann Clarisse S.V. Siababa,² Andrew D. Montecillo¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Here, we report on the draft genomes of five putative biosurfactant-producing isolates of *Serratia* sp. from the East, West, and South bays of Laguna de Bay, Philippines. The contigs ranged from 20 to 21, and the genome lengths ranged from 5,175,553 to 5,214,031 bp.

KEYWORDS Laguna de Bay, *Serratia*, biosurfactant, whole-genome sequencing, WGS, draft genomes

Laguna de Bay is the largest inland wetland lake in the Philippines located around Metro Manila and extends to the province of Rizal and Laguna. It is extensively being used for fishing, source of food, and for irrigation. The extent of pollution is, however, very high with diverse sources of pollutants from household, agriculture, and industry; hence, it is expected to have high concentrations of hydrocarbons (1). The presence of biosurfactant-producing microorganisms would be an important factor in bioremediating such polluted bodies of water.

Five putative biosurfactant-producing bacterial strains were isolated from sediments and water samples from various sites (East, West, and South bays) in Laguna de Bay (14.3935° N, 121.1939° E) (2) following the method of Shoeb et al. (3). Briefly, for each sample, 1 g of sediment or 1 mL of lake water samples was diluted 10-fold using maximum recovery diluent (0.85% NaCl and 0.15% peptone). Dilutions were plated on nutrient agar plates and incubated at 30°C for 24 to 48 h. Single-colony isolates were subcultured on the same medium until axenic cultures were obtained and confirmed microscopically. Cell pellets were sent to Macrogen, Inc., South Korea for 16S rRNA gene and whole-genome sequencing. Sample preparation and DNA extraction were performed following the standard procedures of Macrogen Identification Services. The boiling method using Instagene matrix was used to extract the total DNA, followed by PCR amplification and sequencing. For whole-genome sequencing, TruSeq DNA Nano Library Preparation Kit (Illumina, San Diego, CA) and 2 × 150 bp cycle were used in Illumina NovaSeq 6000 (Illumina, San Diego, CA).

Paired-end raw reads were assessed and trimmed using fastp v.0.23.4 (4) with default parameters. *De novo* assembly was done in Unicycler v.0.5.1 (5) with default parameters, and assembly quality was assessed using QUAST v4.6 (6). Contigs with less than 500 bp length were discarded. The draft genome sequences were annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) (7), and taxonomic placement of the isolates was established using the Microbial Genomes Atlas tool (8) by 16S SSU rRNA classification based on RDP 16S rRNA training set No. 19 07/2023 and RDP Naive Bayesian rRNA Classifier Version 2.14 and by calculating the average nucleotide identity (ANI) in comparison to entries in the NCBI TypeMat database (release 2024-08). Assembly completeness was determined in BUSCO v.5.8.0 (9) with bacteria_odb10 lineage data set. Default parameters were used with all tools, unless otherwise specified. GenBank and SRA accession numbers, sequencing, assembly, and

Editor J. Cameron Thrash, University of Southern California, Los Angeles, California, USA

Address correspondence to Nacita B. Lantican, nblantican@up.edu.ph, or Andrew D. Montecillo, admontecillo@up.edu.ph.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 25 November 2024

Accepted 7 February 2025

Published 10 March 2025

Copyright © 2025 Lantican et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 GenBank and Sequencing Read Archives accession numbers, sequencing library and genome assembly statistics, and annotation of five (5) putative biosurfactant-producing isolates from Laguna de Bay, Philippines

Isolate code	Location and sample type	Whole-genome sequence accession	SRA accession number	Genome size (bp)	Gc (%)	Filtered reads used in the assembly	Mean depth (x)	Contigs	N50 (bp)	Predicted total genes, total CDS, and total protein-coding genes ^a	BUSCO completeness (%)	Nearest genome match
EW14	East Bay—water	JBJHGL0000000000	SRR31327912	5,175,553	60	48,466,944	1,414	21	2,618,309	4,920; 4,832; 4,806	95.2	<i>Serratia sarumani</i> K-M0706 (GCF_029962605.1)
EW4	East Bay—water	JBJHKG0000000000	SRR31327911	5,214,031	60	41,897,142	1,213	20	2,886,712	4,982; 4,892; 4,866	95.2	<i>Serratia sarumani</i> K-M0706 (GCF_029962605.1)
SS15	South Bay—sediment	JBJHJGJ0000000000	SRR31327910	5,213,867	60	48,884,526	1,416	20	2,886,724	4,980; 4,892; 4,866	95.2	<i>Serratia sarumani</i> K-M0706 (GCF_029962605.1)
SS15-1	South Bay—sediment	JBJHJGJ0000000000	SRR31327909	5,213,181	60	45,959,424	1,331	21	2,886,712	4,978; 4,890; 4,864	95.2	<i>Serratia sarumani</i> K-M0706 (GCF_029962605.1)
WS11a	West Bay—sediment	JBJHJGH0000000000	SRR31327908	5,213,854	60	47,766,626	1,383	20	2,886,712	4,980; 4,892; 4,866	95.2	<i>Serratia sarumani</i> K-M0706 (GCF_029962605.1)

^aGenomes were annotated using the NCBI PGAP (7).

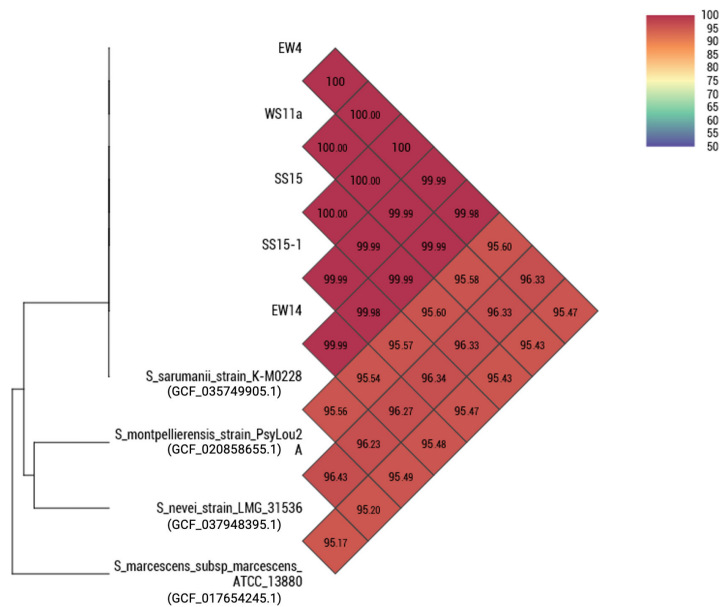


FIG 1 Pairwise average nucleotide identity (ANI) analysis of pucbitative biosurfactant-producing *Serratia* sp. isolates from Laguna de Bay and select reference *Serratia* genomes.

annotation information for each isolate were summarized in Table 1. Digital DNA–DNA hybridization (dDDH) values (formula 2) were calculated against the nearest genome match using the recommended settings in GGDC 4.0 (10). Pairwise ANI and GGDC were calculated in OrthoANI v.0.93.10 (11). The nearest matching species for all genomes is *Serratia sarumanii* (99.9% ANI, 100% dDDH) (Fig. 1).

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support provided by the University of the Philippines System through the Enhanced Creative Work and Research Grant (ECWRG-2022-2-10R) and the EIDR-DOST project titled “Ten years after the Millennium Ecosystem Assessment of Laguna de Bay: Towards Sustainable project for the sediment and water samples.

AUTHOR AFFILIATIONS

¹Microbiology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Banos, Los Banos, Laguna, Philippines

²Genetics and Molecular Biology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Banos, Los Banos, Laguna, Philippines

AUTHOR ORCID_s

Nacita B. Lantican  <http://orcid.org/0000-0002-6105-960X>
 Andrew D. Montecillo  <http://orcid.org/0000-0003-0627-1920>

FUNDING

Funder	Grant(s)	Author(s)
University of the Philippines	ECWRG-2022-2-10R	Nacita B. Lantican

AUTHOR CONTRIBUTIONS

Nacita B. Lantican, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing | Glaezel Angelique T. Barredo, Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing | Albert R. Rosana, Conceptualization, Investigation, Methodology | Ann Clarisse S.V. Siababa, Conceptualization, Investigation, Methodology | Andrew D. Montecillo, Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The sample and sequencing information were uploaded to the NCBI SRA under BioProject accession number [PRJNA1185677](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1185677). The accession numbers for all five SRA files are as follows: [SRR31327912](https://www.ncbi.nlm.nih.gov/sra/SRR31327912), [SRR31327911](https://www.ncbi.nlm.nih.gov/sra/SRR31327911), [SRR31327910](https://www.ncbi.nlm.nih.gov/sra/SRR31327910), [SRR31327909](https://www.ncbi.nlm.nih.gov/sra/SRR31327909), and [SRR31327908](https://www.ncbi.nlm.nih.gov/sra/SRR31327908). This Whole Genome Shotgun project has been deposited in GenBank under accession nos. [JBJHGL000000000](https://www.ncbi.nlm.nih.gov/nuclot/JBJHGL000000000), [JBJHGL000000000](https://www.ncbi.nlm.nih.gov/nuclot/JBJHGL000000000), [JBJHGL000000000](https://www.ncbi.nlm.nih.gov/nuclot/JBJHGL000000000), [JBJHGL000000000](https://www.ncbi.nlm.nih.gov/nuclot/JBJHGL000000000), and [JBJHGL000000000](https://www.ncbi.nlm.nih.gov/nuclot/JBJHGL000000000). The version described in this paper is the first version.

REFERENCES

1. Comiso JC, Espaldon MVQ, Faustino-Eslava DV. 2024. Rediscovering Laguna de Bay: A Vital Natural Resource in Crisis. University of the Philippines Press.
2. Siababa AC. 2022. Characterization and Genome Sequence Analysis of Biosurfactant-Producing Bacteria from Laguna de Bay, Philippines [Unpublished master's thesis], University of the Philippines Los Baños, Philippines
3. Shoeb E, Ahmed N, Akhter J, Badar U, Siddiqui K, Ansari FA, Waqar M, Imtiaz S, Akhtar N, Shaikh Q ul A, Baig R, Butt S, Khan S, Hussain S, Ahmed B, Ansari MA. 2015. Screening and characterization of biosurfactant-producing bacteria isolated from the Arabian Sea coast of Karachi. *Turk J Biol* 39:210–216. <https://doi.org/10.3906/biy-1405-63>
4. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
5. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
6. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
7. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
8. Rodríguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of archaea and bacteria at the whole genome level. *Nucleic Acids Res* 46:W282–W288. <https://doi.org/10.1093/nar/gky467>
9. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
10. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. <https://doi.org/10.1186/1471-2105-14-60>
11. Lee I, Ouk Kim Y, Park SC, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100–1103. <https://doi.org/10.1099/ijsem.0.000760>