

# Complete genome sequence of *Sagittula stellata* strain E-37 reveals a plasmid-encoded type six secretion system

Mary Nia M. Santos,<sup>1</sup> John Raymond B. Salazar,<sup>2</sup> Paul Christian T. Gloria,<sup>3</sup> Alecia N. Septer<sup>4</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** We announce the complete genome sequence of *Sagittula stellata* strain E-37. The hybrid assembly of long and short reads revealed one chromosome and four plasmids. Furthermore, the genome analysis showed that the plasmid-encoded type six secretion system is linked to plasmid replication genes that may be common to Roseobacters.

**KEYWORDS** Roseobacter, type VI secretion system, marine bacterium

Roseobacters are widely distributed in the marine environment, accounting for about 25% of all marine bacteria (1). Aquaculture applications of several strains were demonstrated to be safe and effective against fish pathogens (2). The strain E-37 is a promising biocontrol agent candidate because it is non-pathogenic, is genetically tractable, and can tolerate a wide range of salinity, pH, and temperature (3). Here, we conducted whole genome sequencing to better characterize E-37.

The E-37 strain was isolated from the coastal waters in Georgia, USA (3) and deposited at the American Type Culture Collection (ATCC 700073<sup>T</sup>). Since the E-37 genome is only available as a contig level assembly (GCA\_000169415.1), Illumina and Oxford Nanopore Technologies (ONT) sequencing were used to assemble it into the complete chromosomal level. The E-37 was plated on Seawater Tryptone (SWT) agar from a frozen glycerol stock; then, a colony was grown in SWT broth and incubated overnight at 29°C for DNA extraction using Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research). Illumina sequencing libraries were prepared using the tagmentation-based and PCR-based Illumina DNA Prep kit and custom IDT 10 bp unique dual indices (UDI) with a target insert size of 320 bp. No additional DNA fragmentation or size selection steps were performed. Illumina sequencing was performed on an Illumina NovaSeq X Plus, producing 2 × 151 bp paired end reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert1v4.1.5 (4). The Illumina reads were quality-checked using FastQC v0.11.0 (5) and trimmed using Trimmomatic v0.36.6 (6).

The genomic DNA was extracted using ZymoBIOMICS DNA Miniprep Kit (Zymo Research) from the bacterial cell pellet prior to ONT sequencing. The library was constructed using the Ligation Sequencing Kit (SQK-LSK109) and Native Barcoding Expansion 96 (EXP-NBD196), followed by sequencing on a MinION with a flow cell R9.4.1 (FLO-MIN106D). Base calling was performed using Guppy v6.4.2 with the high-accuracy model. A total of 117,963 reads were generated and quality-trimmed using Porechop v0.2.4 (7). Both long- and short-read sequencing were performed at SeqCenter (Pittsburgh, PA).

Sequencing yielded 508.2 Mb and 606.9 Mb of raw reads from Illumina and ONT platforms, respectively. Unicycler v0.5.0 (8) was used for hybrid assembly, followed by Pilon v1.20.1 (9) for polishing. Unicycler searched for *dnaA* and *repA* alleles in each replicon using TBLASTN to rotate the genome. The genome consisted of one chromosome and four plasmids, all of which are circular contigs and have a total length of

**Editor** Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Mary Nia M. Santos, nia.santos@nfrdi.da.gov.ph, or Alecia N. Septer, asepter@email.unc.edu.

The authors declare no conflict of interest.

See the funding table on p. 2.

**Received** 12 May 2024

**Accepted** 2 October 2024

**Published** 22 October 2024

Copyright © 2024 Santos 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/).

5,290,610 bp and a GC content of 65%. Based on QUAST v5.2.0 (10), the genome yielded 648.1× coverage. The assembly was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v.6.5 (11), which identified 5,257 coding sequences, 13 rRNA genes, 132 tRNA genes, and four ncRNA genes. BUSCO v5.4.5 (12) indicated that the assembly has a completeness score of 100%. The genome encoded type I, II, IV, and VI secretion systems with either complete or fragmented gene members. Interestingly, T6SS is encoded on a plasmid, and the cluster is adjacent to plasmid replication genes, supporting previous findings that T6SSs are prevalent on plasmids in alpha-proteobacteria (13).

## ACKNOWLEDGMENTS

This work was supported by research grants from the National Institutes of Health (R35 GM137886) awarded to ANS and Fulbright-Philippine Agriculture Scholarship Program for Advanced Research in the United States under the Fulbright Program sponsored by the Philippine Department of Agriculture, the US Department of State, Bureau of Educational and Cultural Affairs, and the Philippine-American Educational Foundation (PAEF) awarded to MNMS. We are also grateful to Dr. Mary Ann Moran, University of Georgia, for providing the type culture of E-37 used in this study.

## AUTHOR AFFILIATIONS

<sup>1</sup>Aquaculture Research and Development Division, National Fisheries Research and Development Institute, Quezon City, Philippines

<sup>2</sup>The Graduate School, University of Santo Tomas, Manila, Philippines

<sup>3</sup>Natural Sciences Research Institute, University of the Philippines, Quezon City, Philippines

<sup>4</sup>Department of Earth, Marine and Environmental Sciences, University of North Carolina, Chapel Hill, North Carolina, USA

## AUTHOR ORCIDs

Mary Nia M. Santos  <http://orcid.org/0000-0001-6544-0339>

Paul Christian T. Gloria  <http://orcid.org/0009-0000-7109-4356>

Alecia N. Septer  <http://orcid.org/0000-0002-9309-2930>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">HHS   National Institutes of Health (NIH)</a>	R35 GM137886	Alecia N. Septer
<a href="#">Philippine-American Educational Foundation (PAEF)</a>		Mary Nia M. Santos

## AUTHOR CONTRIBUTIONS

Mary Nia M. Santos, Conceptualization, Data curation, Funding acquisition, Methodology, Supervision, Validation, Writing – original draft, Writing – review and editing | John Raymond B. Salazar, Data curation, Formal analysis, Investigation, Methodology, Validation | Paul Christian T. Gloria, Formal analysis, Methodology, Validation, Writing – review and editing | Alecia N. Septer, Funding acquisition, Project administration, Writing – review and editing

## DATA AVAILABILITY

This genome project has been registered at NCBI under BioProject accession number [PRJNA1109544](#). The genome is available at Genbank under accession number [GCA\\_039724765.1](#). The raw reads have been deposited at the NCBI Sequence Read

Archive under the accession numbers [SRR28974803](https://www.ncbi.nlm.nih.gov/SRR/summary/summary.cgi?acc=SRR28974803) for the ONT raw reads and [SRR28974804](https://www.ncbi.nlm.nih.gov/SRR/summary/summary.cgi?acc=SRR28974804) for the Illumina raw reads.

## REFERENCES

1. Wagner-Döbler I, Biebl H. 2006. Environmental biology of the marine Roseobacter lineage. *Annu Rev Microbiol* 60:255–280. <https://doi.org/10.1146/annurev.micro.60.080805.142115>
2. Sonnenschein EC, Jimenez G, Castex M, Gram L. 2021. The Roseobacter-group bacterium *Phaeobacter* as a safe probiotic solution for aquaculture. *Appl Environ Microbiol* 87:e02581–02520. <https://doi.org/10.1128/AEM.02581-20>
3. Gonzalez J, Mayer F, Moran M, Hodson R, Whitman W. 1997. *Sagittula stellata* gen. nov., sp. nov., a lignin-transforming bacterium from a coastal environment. *Int J Syst Bacteriol* 47:773–780. <https://doi.org/10.1099/00207713-47-3-773>
4. Illumina. 2023. Customer release notes BCL convert standalone v4.1.5. Available from: [https://support.illumina.com/content/dam/illumina-support/documents/downloads/software/BCLConvert/release-notes/200035225\\_00\\_BCL\\_Convert\\_v4.1.5\\_Customer\\_Release\\_Notes.pdf](https://support.illumina.com/content/dam/illumina-support/documents/downloads/software/BCLConvert/release-notes/200035225_00_BCL_Convert_v4.1.5_Customer_Release_Notes.pdf)
5. Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
7. Wick R. 2017. In GitHub repository. Porechop. <https://github.com/rrwick/Porechop>.
8. 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>
9. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
10. Mikheenko A, Prijibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
11. 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>
12. 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>
13. Morgado S, Vicente AC. 2022. Diversity and distribution of type VI secretion system gene clusters in bacterial plasmids. *Sci Rep* 12:8249. <https://doi.org/10.1038/s41598-022-12382-3>