



Genome Update for *Pseudomonas fluorescens* Isolate SBW25

 Carsten Fortmann-Grote,^a  Eric Hugoson,^{a*}  Joanna Summers,^a  Loukas Theodosiou,^a  Paul B. Rainey^{a,b}

^aDepartment of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön, Germany

^bLaboratoire Biophysique et Évolution, Institut Chimie Biologie Innovation, ESPCI Paris, Université Paris Sciences et Lettres, CNRS, Paris, France

ABSTRACT We report a genome update for *Pseudomonas fluorescens* isolate SBW25. The updated genome assembly, which was derived from the original isolate, is based on PacBio long-read sequence data. It shows three minor differences, compared with the previously published genome sequence. Original annotations were merged with recent automated annotations to preserve information.

Pseudomonas fluorescens is a Gram-negative bacterium that is commonly found in soil, in water, and on plant surfaces (1). Strain SBW25 (taxonomy identification number [NCBIT:taxid:216595](#)) is a model organism in evolution, population biology, microbe-plant interactions, and genetics. The *Pseudomonas fluorescens* SBW25 genome was originally sequenced at the Wellcome Sanger Institute (2). Sample collection and isolation are detailed in references 3 and 4. The Wellcome Sanger Institute assembled and annotated the 6,722,539-bp-long genome (2). The annotation has repeatedly been updated with automated annotation pipelines such as Prokka (5) and PGAP (6).

Our updated *Pseudomonas fluorescens* SBW25 genome is a *de novo* hybrid assembly of Pacific Biosciences (PacBio) Sequel single-molecule real-time (SMRT) long reads and Illumina NextSeq 500 short reads for our stock sample, MPB13884. MPB13884 was obtained by passaging the original strain PBR340, which was held at -80°C , through three consecutive single-colony-derived overnight 6-mL cultures grown at 28°C , representing approximately 100 to 150 generations.

High-density genomic DNA was extracted using the Nanobind CBB Big DNA kit (Circulomics). The library was prepared with the PacBio template preparation kit v1.0 (7). Initial shearing was performed with a MegaRuptor device, with size selection at 15 kbp. The raw data comprised 435,628 reads, with a mean length of 6,581 bp (coverage, $425\times$).

The raw PacBio reads were filtered using Filtlong (8) (options: `-min_length 1000 -keep_percent 90`). We used three different assemblers, i.e., Flye v2.5 (9) (options: `-genome-size 6722539 -threads 15 -plasmids`), miniasm v0.3 followed by minimap2 v0.13 (10, 11) (both with option: `-threads 15`), and Raven v1.8 (12) (option: `-threads 15`). We produced the consensus assembly using Tricycler v0.4.1 (10, 13) (options: `cluster -threads 15` followed by `reconcile -threads 15 cluster_001`).

Illumina short-read sequencing was performed at the Max Planck Institute for Evolutionary Biology on the NextSeq 500 platform. DNA was extracted with the Wizard genomic DNA purification kit A1120 (Promega). Library preparation with the Nextera XT kit (Illumina) followed the protocol described previously (14). The short-read data (49,451,604 paired-end reads between 35 and 151 bp in length, with 71% of reads with 150-bp length [coverage, $54\times$]) were preprocessed with fastp v0.20.1 (15) (default options for quality trimming and removal of Illumina adaptors). Pilon v1.23 (16) (default options) produced the hybrid assembly.

We aligned the old and new assemblies with MAFFT v1.4.0 (17) (default options). Compared to the old assembly, the new assembly is 139 bp shorter (6,722,400 bp versus 6,722,539 bp); the difference is caused by one 356-bp deletion (between position 3694247 and position 3694603) and two insertions, one of 3 bp (position

Editor Catherine Putonti, Loyola University Chicago

Copyright © 2023 Fortmann-Grote et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Carsten Fortmann-Grote, carsten.fortmann-grote@evolbio.mpg.de.

*Present address: Eric Hugoson, Laboratoire National de Santé, Dudelange, Luxembourg. The authors declare no conflict of interest.

Received 29 July 2022

Accepted 22 December 2022

TABLE 1 Genome assembly and annotation accession numbers for *Pseudomonas fluorescens* SBW25

Entity	Accession no.
Assembly and annotation	OV986001
PacBio raw reads	ERR8265873
Illumina raw reads	ERR9795577
BioSample	ERS379752
BioProject	PRJEA31229

3448192 to position 3448195) and the other of 214 bp (position 895392 to position 895606).

To annotate the new assembly, we applied the following procedure. Public annotations for *Pseudomonas fluorescens* SBW25 were transferred from RefSeq (GenBank assembly accession number [GCF_000009225.2](#)), GenBank (accession number [GCA_000009225.1](#)), ENA (accession number [GCA_000009225](#)), and Sanger Institute (accession number [AM181176](#)) databases to the new assembly using Geneious Prime v2020.2.4 (<https://www.geneious.com>), applying the protocol described in reference 18 (using the method presented in exercise 3). These transferred annotations were loaded as relational database tables using gffutils v0.10.1 (19). Annotated regions with identical start, end, strand, and feature type were merged into single features, maintaining all distinct attributes. The merged table was serialized into gff3 format using Biopython v1.79 (20), converted into EMBL format using EMBLmyGFF3 v1.3 (21) (options: `–topology circular –molecule_type genomic DNA –transl_table 11 –species Pseudomonas fluorescens –strain SBW25 –locus_tag PFLU –project_id PRJEA31229 –data_class WGS, –x PRO –use_attribute_value_as_locus_tag locus_tag –interleave_genes –version 1`), and submitted to the ENA.

Data availability. Table 1 lists all relevant accession numbers. All scripts to produce the submitted assembly and annotation are available at https://gitlab.gwdg.de/mpievollbio-scicomp/sbw25_newgenome/-/tree/assembly_scripts.

ACKNOWLEDGMENTS

We thank Dave Rogers for his valuable assistance. We acknowledge generous core support from the Max Planck Society. We thankfully acknowledge the Long-Read Team of the DRESDEN Concept Genome Center, part of the Max Planck Institute of Molecular Cell Biology and Genetics, the technology platform of the Center for Molecular and Cellular Bioengineering at the TU Dresden (supported by DFG grant INST 269/768-1), and the Sequencing Core Facility at the Max Planck Institute for Evolutionary Biology.

REFERENCES

- Bailey MJ, Lilley AK, Thompson IP, Rainey PB, Ellis RJ. 1995. Site directed chromosomal marking of a fluorescent pseudomonad isolated from the phytosphere of sugar beet: stability and potential for marker gene transfer. *Mol Ecol* 4:755–763. <https://doi.org/10.1111/j.1365-294x.1995.tb00276.x>.
- Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang X-X, Moon CD, Gehrig SM, Godfrey SAC, Knight CG, Malone JG, Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hotherhall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB, Thomson NR. 2009. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol* 10:R51. <https://doi.org/10.1186/gb-2009-10-5-r51>.
- Thompson IP, Bailey MJ, Fenlon JS, Fermor TR, Lilley AK, Lynch JM, McCormack PJ, McQuilken MP, Purdy KJ, Rainey PB, Whipps JM. 1993. Quantitative and qualitative seasonal-changes in the microbial community from the phyllosphere of sugar-beet (*Beta vulgaris*). *Plant Soil* 150:177–191. <https://doi.org/10.1007/BF00013015>.
- Thompson IP, Bailey MJ, Ellis RJ, Lilley AK, McCormack PJ, Purdy KJ, Rainey PB. 1995. Short-term community dynamics in the phyllosphere microbiology of field-grown sugar-beet. *FEMS Microbiol Ecol* 16:205–212. <https://doi.org/10.1111/j.1574-6941.1995.tb00284.x>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvermin 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>.
- Pacific Biosciences. 2018. Multiplexed microbial SMRTbell libraries for the PacBio Sequel system protocol, version 2. Part number 101-489-000. Pacific Biosciences, Menlo Park, CA.
- Wick R. 2021. Filtlong. <https://github.com/rwick/Filtlong>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Wick RR, Holt KE. 2019. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* 8:2138. <https://doi.org/10.12688/f1000research.21782.4>.
- Wick R. 2020. Minipolish. Zenodo <https://doi.org/10.5281/ZENODO.3752204>.

12. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 1:332–336. <https://doi.org/10.1038/s43588-021-00073-4>.
13. Wick RR. 2021. Trycycler. Zenodo <https://doi.org/10.5281/ZENODO.5769082>.
14. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. <https://doi.org/10.1371/journal.pone.0128036>.
15. 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>.
16. 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>.
17. Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
18. Geneious. 2022. Transferring annotations between sequences tutorial. <https://www.geneious.com/tutorials/transferring-annotations>.
19. Dale R. 2022. gffutils. <http://daler.github.io/gffutils>.
20. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
21. Norling M, Jareborg N, Dainat J. 2018. EMBLmyGFF3: a converter facilitating genome annotation submission to European Nucleotide Archive. *BMC Res Notes* 11:584. <https://doi.org/10.1186/s13104-018-3686-x>.