




Draft Genome Assembly of *Rhodobacter sphaeroides* 2.4.1 Substrain H2 from Nanopore Data

 Robert Maximilian Leidenfrost,^a Nadine Wappler,^a  Röbbbe Wünschiers^a

^aDepartment of Biotechnology and Chemistry, Mittweida University of Applied Sciences, Mittweida, Germany

ABSTRACT *Rhodobacter sphaeroides* is a purple bacterium with complex genomic architecture. Here, a draft genome is reported for *R. sphaeroides* strain 2.4.1 substrain H2, which was generated exclusively from Nanopore sequencing data.

Rhodobacter sphaeroides 2.4.1 belongs to the phylogenetically distinct α -3 group of alphaproteobacteria and is capable of facultative photosynthesis. It has been well studied as a photosynthetic system and is considered the *R. sphaeroides* type strain. It was originally described by van Niel in 1944 (1), a near-complete genome was published by Mackenzie et al. in 2001 (2), and the sequence was revised by Kontur et al. in 2012 (3) (see BioProject accession number [PRJNA56](#) in Fig. 1). Here, we report the Nanopore sequencing based *de novo* assembly for strain *R. sphaeroides* 2.4.1 substrain H2 (4), which is of particular interest and under current investigation within the context of photofermentative hydrogen production. It was acquired from TU Dresden and evolved serendipitously in the laboratory from the type strain (ATH 2.4.1; also named ATCC 17023, IAM 14237, and NCIB 8253), which was originally obtained as DSM 158 from the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). For reference, we deposited the Nanopore raw read sequences of DSM 158 under SRA accession number [SRX7341766](#).

Cultivation was performed in medium 112 (Van Niel's yeast medium) at 33°C. Total DNA was extracted using the MasterPure total DNA purification kit (Epicentre).

Genomic DNA was sheared using g-TUBEs (Covaris) according to the protocol, purity was assessed using a NanoVue spectrophotometer (GE Healthcare), and the quantity was determined using a Qubit fluorometer with a high-sensitivity assay kit (Invitrogen). The library was prepared for 1D sequencing with an SQK-LSK108 kit (Oxford Nanopore Technologies [ONT]) and barcoded as part of a multiplexed sequencing run (EXP-NBD103; ONT). Sequencing was performed with an R9.4.1 flow cell (FLO-MIN106; ONT).

Base calling was performed with Guppy v3.0.3 (ONT), and results were further processed with Porechop v0.2.3 (<https://github.com/rrwick/Porechop>) (parameters: `-b --barcode_threshold 85 --require_two_barcodes --discard_middle`), i.e., demultiplexed and concurrently adapter trimmed. Read data quality was assessed with NanoPlot v1.27.0 (5). Assemblies were computed with Flye v2.5 (6) and consensus polished three times with Racon v1.4.3 (7); each contig was rotated and then signal-level polished twice with Nanopolish v0.11.2 (8). Starting points were adjusted to match those reported by Kontur et al. (3). Alignments were performed with minimap2 v2.11 (9). Polished assemblies were visually checked with Gepard v1.40 (10) and automatically annotated using PGAP upon submission to the NCBI database (11, 12). All processing was performed with default software parameters, unless otherwise specified.

Sequencing of 2.4.1 substrain H2 yielded 434 Mb of base-called and demultiplexed raw data (total number of reads, 35,752; mean length, 12,136 bp; N_{50} , 23,716 bp), which were not further trimmed. The assembly is presented in six circular contigs, with an

Citation Leidenfrost RM, Wappler N, Wünschiers R. 2020. Draft genome assembly of *Rhodobacter sphaeroides* 2.4.1 substrain H2 from Nanopore data. Microbiol Resour Announc 9:e00414-20. <https://doi.org/10.1128/MRA.00414-20>.

Editor J. Cameron Thrash, University of Southern California

Copyright © 2020 Leidenfrost et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Röbbbe Wünschiers, wuenschi@hs-mittweida.de.

Received 14 April 2020

Accepted 22 June 2020

Published 16 July 2020

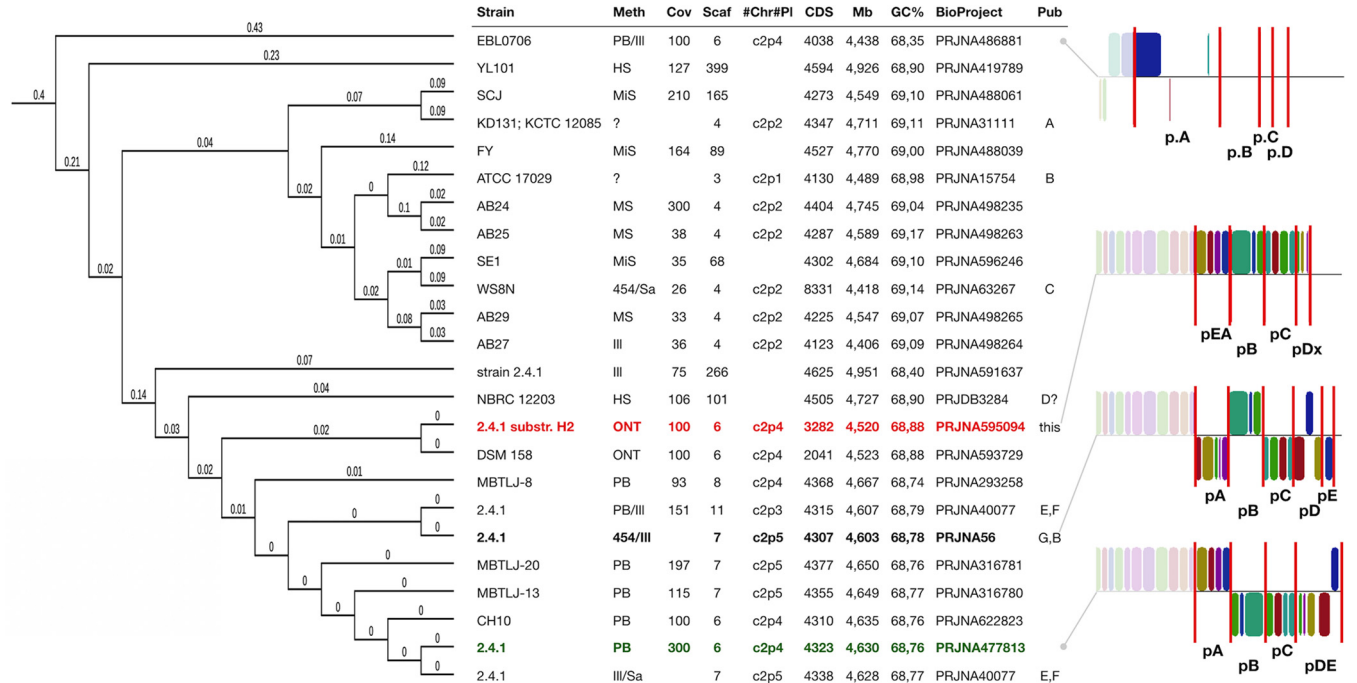


FIG 1 Compilation of available *R. sphaeroides* genome data. The BLAST tree display was created with NCBI Tree Viewer v1.17.6. It shows the autogenerated NCBI Genome Tree report (Tree 509) from genetic distances calculated from the aligned genome sequences using the Jukes-Cantor (16) substitution model. The tree was built from this distance matrix using FastME (17). BioProject number PRJNA376580 was used as an outgroup. The branch lengths are not to scale, and the numbers present percent genetic variation. In the table, substrain H2 and the former (PRJNA56) and current (PRJNA477813) reference strain 2.4.1 genomes are in bold. For the architecture diagrams, full-genome alignments for four strains were calculated with progressiveMauve v2.4.0 (18). Colored blocks represent homologous regions. Blocks below the center line indicate regions that align in the inverse orientation. Original plasmid labels were retained. While strain EBL0706 differed substantially, the rest were identical with respect to the chromosomes but differed in plasmid number and architecture. Meth, sequencing method; Cov, coverage; Scaf, scaffolds; #Chr#Pl, numbers of chromosomes and plasmids; CDS, coding sequences; Mb, genome size; GC%, GC content; Pub, related publication (A [19], B [20], C [21], D [22], E [23], F [24], G [3]); PB, PacBio single-molecule sequencing; III, Illumina; HS, Illumina HiSeq; MS, Illumina MiSeq; MiS, Illumina MiniSeq; 454, 454 GS FLX Titanium; ONT, ONT MinION; Sa, Sanger sequencing.

overall size of 4,519,621 bp. The contigs correspond to two chromosomes, 3.2 Mb and 0.9 Mb in size, and four plasmids, ~124 kb, ~114 kb, ~105 kb, and ~44 kb in size (Fig. 1). The sequence possesses an average coverage of 100×, an N_{50} value of 3,188,040 bp, and a mean GC content of 68.9%.

Notably, PGAP reports a large number of pseudogenes. These are reportedly caused by indels leading to frameshifts, which are known to be an error source in Nanopore sequencing-only assemblies (13) and must be taken into account for downstream purposes, such as annotation. While the plasmid architecture is substantially divergent from the sequence reported by Kontur et al. (3) (accession numbers ASM1290v2 and PRJNA56) (Fig. 1), such divergence may also be observed in the literature (14, 15) and for more recently published genome assemblies (accession numbers ASM332471v1/PRJNA477813 and ASM342926v1/PRJNA486881) (Fig. 1).

Data availability. The draft assembly of *R. sphaeroides* 2.4.1 substrain H2 has been deposited in NCBI GenBank under accession number ASM979766v1. The SRA deposit is available under accession number SRX7352322. The Nanopore raw read sequences of DSM 158 were deposited under SRA accession number SRX7341766.

ACKNOWLEDGMENTS

R.M.L. and N.W. acknowledge support from the European Social Fund (grants 100316182 and 100235472, respectively). R.W. received funding from the Saxon State Ministry of Science and Art and the Saxony5 Initiative.

We thank Karsten Helbig from TU Dresden (Dresden, Germany), Department of Bioprocess Engineering, for supplying *R. sphaeroides* 2.4.1 substrain H2.

REFERENCES

- van Niel CB. 1944. The culture, general physiology, morphology, and classification of the non-sulfur purple and brown bacteria. *Bacteriol Rev* 8:1–118. <https://doi.org/10.1128/MMBR.8.1.1-118.1944>.
- Mackenzie C, Choudhary M, Larimer FW, Predki PF, Stilwagen S, Armitage JP, Barber RD, Donohue TJ, Hosler JP, Newman JE, Shapleigh JP, Sockett RE, Zeilstra-Ryalls J, Kaplan S. 2001. The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. *Photosynth Res* 70:19–41. <https://doi.org/10.1023/A:1013831823701>.
- Kontur WS, Schackwitz WS, Ivanova N, Martin J, Labutti K, Deshpande S, Tice HN, Pennacchio C, Sodergren E, Weinstock GM, Noguera DR, Donohue TJ. 2012. Revised sequence and annotation of the *Rhodobacter sphaeroides* 2.4.1 genome. *J Bacteriol* 194:7016–7017. <https://doi.org/10.1128/JB.01214-12>.
- Krujatz F, Härtel P, Helbig K, Haufe N, Thierfelder S, Bley T, Weber J. 2015. Hydrogen production by *Rhodobacter sphaeroides* DSM 158 under intense irradiation. *Bioresour Technol* 175:82–90. <https://doi.org/10.1016/j.biortech.2014.10.061>.
- de Coster W, D'Hert S, Schultz DT, Cruys M, van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- 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>.
- Vaser R, Sović I, Nagarajan N, Sikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
- Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only Nanopore sequencing data. *Nat Methods* 12:733–735. <https://doi.org/10.1038/nmeth.3444>.
- Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23:1026–1028. <https://doi.org/10.1093/bioinformatics/btm039>.
- 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>.
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>.
- Taylor TL, Volkening JD, DeJesus E, Simmons M, Dimitrov KM, Tillman GE, Suarez DL, Afonso CL. 2019. Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read Nanopore technology. *Sci Rep* 9:16350. <https://doi.org/10.1038/s41598-019-52424-x>.
- Saunders VA, Saunders JR, Bennett PM. 1976. Extrachromosomal deoxyribonucleic acid in wild-type and photosynthetically incompetent strains of *Rhodospseudomonas sphaeroides*. *J Bacteriol* 125:1180–1187. <https://doi.org/10.1128/JB.125.3.1180-1187.1976>.
- Fornari CS, Watkins M, Kaplan S. 1984. Plasmid distribution and analyses in *Rhodospseudomonas sphaeroides*. *Plasmid* 11:39–47. [https://doi.org/10.1016/0147-619x\(84\)90005-2](https://doi.org/10.1016/0147-619x(84)90005-2).
- Jukes T, Cantor C. 1969. Evolution of protein molecules, p 21–132. *In* Munro HN, Allison J (ed), *Mammalian protein metabolism*, vol 3. Academic Press, New York, NY.
- Desper R, Gascuel O. 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol* 9:687–705. <https://doi.org/10.1089/106652702761034136>.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
- Lim S-K, Kim SJ, Cha SH, Oh Y-K, Rhee H-J, Kim M-S, Lee JK. 2009. Complete genome sequence of *Rhodobacter sphaeroides* KD131. *J Bacteriol* 191:1118–1119. <https://doi.org/10.1128/JB.01565-08>.
- Choudhary M, Zanhua X, Fu YX, Kaplan S. 2007. Genome analyses of three strains of *Rhodobacter sphaeroides*: evidence of rapid evolution of chromosome II. *J Bacteriol* 189:1914–1921. <https://doi.org/10.1128/JB.01498-06>.
- Porter SL, Wilkinson DA, Byles ED, Wadhams GH, Taylor S, Saunders NJ, Armitage JP. 2011. Genome sequence of *Rhodobacter sphaeroides* strain WS8N. *J Bacteriol* 193:4027–4028. <https://doi.org/10.1128/JB.05257-11>.
- Reyes F, Gavira M, Castillo F, Moreno-Vivián C. 1998. Periplasmic nitrate-reducing system of the phototrophic bacterium *Rhodobacter sphaeroides* DSM 158: transcriptional and mutational analysis of the *napKEDABC* gene cluster. *Biochem J* 331:897–904. <https://doi.org/10.1042/bj3310897>.
- Ribeiro FJ, Przybylski D, Yin S, Sharpe T, Gnerre S, Abouelleil A, Berlin AM, Montmayeur A, Shea TP, Walker BJ, Young SK, Russ C, Nusbaum C, MacCallum I, Jaffe DB. 2012. Finished bacterial genomes from shotgun sequence data. *Genome Res* 22:2270–2277. <https://doi.org/10.1101/gr.141515.112>.
- MacCallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, Williams L, Young S, Nusbaum C, Jaffe DB. 2009. ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol* 10:R103. <https://doi.org/10.1186/gb-2009-10-10-r103>.