JB Journals.ASM.org

GENOME ANNOUNCEMENT

Revised Sequence and Annotation of the *Rhodobacter sphaeroides* 2.4.1 Genome

Wayne S. Kontur,[a,b] Wendy S. Schackwitz,[c] Natalia Ivanova,[c] Joel Martin,[c] Kurt LaButti,[c] Shweta Deshpande,[c] Hope N. Tice,[c] Christa Pennacchio,[c] Erica Sodergren,[d] George M. Weinstock,[d] Daniel R. Noguera,[b,e] and Timothy J. Donohue[a,b]

Department of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin, USA[a]; DOE Great Lakes Bioenergy Research Center, Madison, Wisconsin, USA[b]; DOE Joint Genome Institute, Walnut Creek, California, USA[c]; The Genome Institute, Washington University, St. Louis, Missouri, USA[d]; and Department of Civil and Environmental Engineering, University of Wisconsin—Madison, Madison, Wisconsin, USA[e]

The DNA sequences of chromosomes I and II of *Rhodobacter sphaeroides* strain 2.4.1 have been revised, and the annotation of the entire genomic sequence, including both chromosomes and the five plasmids, has been updated. Errors in the originally published sequence have been corrected, and ~11% of the coding regions in the original sequence have been affected by the revised annotation.

Rhodobacter sphaeroides is a purple nonsulfur photosynthetic bacterium belonging to the alpha class of proteobacteria. *R. sphaeroides* is metabolically diverse, capable of growing photo- or chemoheterotrophically or -autotrophically and of fixing dinitrogen.

The *R. sphaeroides* 2.4.1 genome was originally sequenced in 2001 (3) (www.rhodobacter.org) and consists of two chromosomes (I and II) and five plasmids (A, B, C, D, and E). Recently, after sequencing the genome of a mutant strain, we identified a number of potential errors in the published 2.4.1 sequence. Here, we report revised sequences of chromosomes I and II of *Rhodobacter sphaeroides* 2.4.1 and a revised annotation of the entire genomic sequence, including the five plasmids.

Genomic DNA was randomly sheared into ~200-bp fragments. The library created from these fragments was sequenced on an Illumina GAIIx system. The resulting 36-bp paired-end reads were aligned to the original 2.4.1 genomic sequence using the software program maq-0.7.1. (1) (www.maq.sourceforge.net). An iterative process utilizing maq-0.7.1 and manual inspection was used to identify single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) between the reads and the reference sequence. We also used whole-genome sequencing data from a 454 FLX Titanium platform and a combination of PCR and Sanger sequencing to verify areas of low read coverage in the Illumina data.

Compared to the original genomic sequence, there are 106 SNPs in the revised genomic sequence. There are 37 insertions, comprising 141 bp total (the largest spanning 46 bp), and 73 deletions, comprising 200 bp total (the largest spanning 45 bp).

The entire genomic sequence (including the five plasmids, which retained their original sequences) was annotated using the JGI microbial genome automated annotation pipeline (5), followed by manual curation using the GenePRIMP software program (6). The predicted coding sequences (CDSs) were translated and used to search the NCBI nonredundant database and the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction and functional annotation were performed within the IMG-ER platform (4).

The revised sequence contains 4,515 predicted CDSs (versus 4,383 in the original sequence). Of the CDSs in the original sequence, 89% remain unmodified. Of the modified CDSs, 173 were extended and 202 were truncated in length; these modified CDSs retained their original locus tags. In 42 cases, two or more CDSs from the original sequence were combined into a single CDS, which retained the locus tag of the original CDS furthest upstream. Twenty-three CDSs were replaced by new CDSs (with new locus tags) on the opposite DNA strand, and seven CDSs were eliminated. All CDSs new to the revised genomic sequence have been given locus tags beginning with RSP_7500. All modified CDSs contain a description of the modification as a note in their GenBank and IMG entries. Most modifications resulted in CDSs that more closely match those of other sequenced *R. sphaeroides* strains (2, 7).

Nucleotide sequence accession numbers. The newly sequenced and annotated genomic chromosomes and the newly annotated genomic plasmids have been deposited in GenBank under the same accession numbers as those for the original *R. sphaeroides* 2.4.1 genome (CP000143 to CP000147, DQ232586, and DQ232587).

ACKNOWLEDGMENTS

We thank Alexandra Linz and the DNA Sequencing Facility at the University of Wisconsin—Madison Biotechnology Center for help with Sanger sequencing of regions of the genome.

This work was funded in part by the U.S. Department of Energy, Office of Science's Biological and Environmental Research program (DE-FG02-07ER64495) and the Great Lakes Bioenergy Research Center (DE-FC02-07ER64494). The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. G.M.W. acknowledges support from NIH U54HG004968.

REFERENCES

1. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**:1851–1858.

Received 11 July 2012 **Accepted** 11 October 2012

Address correspondence to Timothy J. Donohue, tdonohue@bact.wisc.edu.

W.S.K., W.S.S., and N.I. contributed equally to this work.

doi:10.1128/JB.01214-12

2. **Lim S, et al.** 2009. Complete genome sequence of *Rhodobacter sphaeroides* KD131. J. Bacteriol. **191**:1118–1119.
3. **Mackenzie C, et al.** 2001. The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. Photosynth. Res. **70**:19–41.
4. **Markowitz VM, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.
5. **Mavromatis K, et al.** 2009. Standard operating procedure for the annotations of microbial genomes by the production genomic facility of the DOE JGI. Stand. Genomic Sci. **1**:63–67.
6. **Pati A, et al.** 2010. GenePRIMP: a gene prediction improvement pipeline for microbial genomes. Nat. Methods **7**:455–457.
7. **Porter SL, et al.** 2011. Genome sequence of *Rhodobacter sphaeroides* strain WS8N. J. Bacteriol. **193**:4027–4028.