Vol. 192, No. 13

GENOME ANNOUNCEMENTS

Complete Genome Sequence of the Photosynthetic Purple Nonsulfur Bacterium *Rhodobacter capsulatus* SB 1003[▽]

Hynek Strnad,[1] Alla Lapidus,[2] Jan Paces,[1] Pavel Ulbrich,[3] Cestmir Vlcek,[1]
Vaclav Paces,[1] and Robert Haselkorn[4]*

*Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 14220 Prague, Czech Republic[1];
DOE-Joint Genome Institute, Walnut Creek, California 94598[2]; Institute of Chemical Technology,
16628 Prague, Czech Republic[3]; and Department of Molecular Genetics and Cell Biology,
University of Chicago, Chicago Illinois 60637[4]*

Rhodobacter capsulatus **SB 1003 belongs to the group of purple nonsulfur bacteria. Its genome consists of a 3.7-Mb chromosome and a 133-kb plasmid. The genome encodes genes for photosynthesis, nitrogen fixation, utilization of xenobiotic organic substrates, and synthesis of polyhydroxyalkanoates. These features made it a favorite research tool for studying these processes. Here we report its complete genome sequence.**

The genome of *Rhodobacter capsulatus* SB 1003 consists of a single chromosome containing 3,738,958 bp and a circular plasmid of 132,962 bp. Both the chromosome and the plasmid have relatively high GC content (66.6%) that made DNA sequencing difficult. The project was started by constructing a library of overlapping cosmids covering the whole genome (4, 5). Sequencing was performed by using standard Sanger approaches, and the average sequencing depth was 7. The complete sequence was analyzed by using Critica (1) and Glimmer (3) for the protein-coding genes, tRNAscan (11) and Aragorn (10) for the tRNA and tmRNA genes, and RNAmmer (8) for the rRNA genes. The functions of the predicted protein-coding genes were annotated by comparison with the UniRef90 (14), NCBI-NR (2), COG (15), and KEGG (7) databases. The annotation results were verified using Artemis (13) with dicodon use plots (12).

We found 3,531 open reading frames (ORFs) in the chromosome and 154 ORFs in the plasmid. There are 4 rRNA operons and 53 tRNA genes for all 20 amino acids and one tmRNA gene. The coding density of the *R. capsulatus* genome is 91%. Functions were assigned to 3,100 ORFs (84.1%). Six hundred ten ORFs (16.6%) represent genes with some similarity to hypothetical genes in databases. The remaining ORFs had no homologues in the databases (e-value $< 1 \times 10^{-10}$). There are 42 putative intact or mutated transposase genes and 237 phage-related genes. An important and specific feature of *R. capsulatus* is the defective phage called *gene transfer agent* (GTA) that provides a useful tool for genetic analysis (9). The distribution of genes among the COG metabolic functional classes and detailed descriptions of the metabolic pathways can

be found at http://rhodo.img.cas.cz. Based on the GC skew analysis and the orientation of transcription, the origin of replication was localized around 160 kb.

Five complete restriction-modification (RM) systems were identified in the genome. Three of them are type I (RcaSBIP, RcaSBIIIP, and RcaSBIV), one is type III (RcaSBIIP), and one is a type IV system (RcaSBMcrBCP). These RM systems are located mainly in regions of exogenic origin. Several genes encoding these restriction and modification enzymes were isolated and characterized. RcaSBIV recognizes the sequence AGAN7RTAG (H. Strnad and V. Paces, unpublished). Eight CRISPR/Cas systems (6), of which two comprise a considerable number of repetitions (41 and 9), were also found. The others comprise three to five repetitions.

Nucleotide sequence accession numbers. The nucleotide sequences were deposited in GenBank with accession numbers CP001312 (chromosome) and CP001313 (plasmid pRCB133). Complete analysis of the genome is accessible at http://rhodo.img.cas.cz.

REFERENCES

1. **Badger, J., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
2. **Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler.** 2008. GenBank. Nucleic Acids Res. **36:**D25–D30.
3. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
4. **Fonstein, M., and R. Haselkorn.** 1993. Chromosomal structure of *Rhodobacter capsulatus* strain SB1003: cosmid encyclopedia and high-resolution physical and genetic map. Proc. Natl. Acad. Sci. U. S. A. **90:**2522–2526.
5. **Haselkorn, R., A. Lapidus, Y. Kogan, C. Vlcek, J. Paces, V. Paces, P. Ulbrich, T. Pecenkova, D. Rebrekov, A. Milgram, M. Mazur, R. Cox, N. Kyprides, N. Ivanova, V. Kapatral, T. Los, A. Lykidis, N. Mikhailova, G. Reznik, O. Vasieva, and M. Fonstein.** 2001. The *Rhodobacter capsulatus* genome. Photosynth. Res. **70:**43–52.

* Corresponding author. Mailing address: Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago IL 60637. Phone: (773) 702-1069. Fax: (773) 702-2853. E-mail: rh01@uchicago.edu.

6. **Horvath, P., and R. Barrangou.** 2010. CRISPR/Cas, the immune system of Bacteria and Archaea. Science **327:**167–170.

7. **Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi.** 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. **36:**D480–D484.

8. **Lagesen, K., P. F. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery.** 2007. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35:**3100–3108.

9. **Lang, A. S., T. A. Taylor, and J. T. Beatty.** 2002. Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of Rhodobacter capsulatus. J. Mol. Evol. **55:**534–543.

10. **Laslett, D., and B. Canback.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16.

11. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.

12. **Paces, J., and V. Paces.** 2002. DicodonUse: the program for dicodon bias visualization in prokaryotes. Folia Biol. (Praha). **48:**246–249.

13. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–955.

14. **Suzek, B. E., H. Huang, P. McGarvey, R. Mazumder, and C. H. Wu.** 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics **23:**1282–1288.

15. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.