genomeA
Journals.ASM.org

Draft Genome Sequence of an Alphaproteobacterium Associated with the Mediterranean Sponge *Oscarella lobularis*

Cyril Jourda,[a] Sébastien Santini,[a] Caroline Rocher,[b] André Le Bivic,[c] Jean-Michel Claverie[a,d]

Aix-Marseille Université, CNRS UMR 7256 (IMM FR 3479), Structural and Genomic Information Laboratory, Marseille, France[a]; Aix-Marseille Université, CNRS UMR 7263, Mediterranean Institute of Marine and Continental Biodiversity and Ecology (IMBE), Station Marine d'Endoume, Marseille, France[b]; Aix-Marseille Université, CNRS UMR 7288, Developmental Biology Institute of Marseille Luminy (IBDM), Marseille, France[c]; Assistance Publique des Hopitaux de Marseille, Marseille, France[d]

While sequencing DNA purified from the homoscleromorph sponge *Oscarella lobularis*, we detected a large number of reads with strong similarity to available alphaproteobacteria gene sequences of family *Rhodobacteraceae*. Here, we present the genome sequence of this putative sponge symbiont that we propose to designate as "*Candidatus* Rhodobacter lobularis."

Received 20 July 2015 **Accepted** 24 July 2015 **Published** 3 September 2015

Citation Jourda C, Santini S, Rocher C, Le Bivic A, Claverie J-M. 2015. Draft genome sequence of an alphaproteobacterium associated with the Mediterranean sponge *Oscarella lobularis*. Genome Announc 3(5):e00977-15. doi:10.1128/genomeA.00977-15.

Address correspondence to Jean-Michel Claverie, claverie@igs.cnrs-mrs.fr.

Sponges (*Porifera* phylum) are metazoans divided into four major classes, including *Homoscleromorpha*, *Demospongiae*, *Hexactinellida*, and *Calcarea*. *Oscarella lobularis* is a *Homoscleromorpha* sponge endemic to the Mediterranean Sea which has several color morphs and host four morphotypes of microbes with dominance of the *Alphaproteobacteria* species (1). Here, we present the draft genome sequence of a new member of the *Rhodobacteraceae* family found in high abundance associated to *Oscarella lobularis* sampled in cave Endoume, Marseille, France.

We initiated the genome sequencing using Illumina technology with DNA-seq paired ends and Nextera mate-pair protocols on a HiSeq2500 sequencer. Low-quality read ends ($Q < 28$), adapter, and cloning vector sequences were trimmed and short remaining sequences (< 100 pb) were removed using Cutadapt (2). The remaining 112 million paired reads were assembled with IDBA-UD (3) and the scaffolding tool from the Platanus assembler (4). Ends and gaps within scaffolds were tentatively closed using GapFiller (5). Extended scaffolds were assembled using CAP3 (6). Read pairs were mapped onto scaffolds using Bowtie (7) and sequencing coverage levels were estimated for each scaffold using SAMtools (8). All 16S sequences were retrieved and taxonomically annotated following homology searches run on the SILVA database (9).

A typical 16S *Rhodobacteraceae*-like sequence was determined at a high coverage rate (1,005×), approximately corresponding to 20 bacterial cells per sponge cell. We propose "*Candidatus* Rhodobacter lobularis" as a name for this new species based on its 16S taxonomic classification and its association with the sponge. The two large scaffolds corresponding to this species were pointed out by their high sequencing coverage (>1,070× in average) and further characterized through a similarity search of their predicted genes (using MetaGeneMark [10]) on the NR database (best BLASTp hits, $E < 10^{-5}$). The resulting 5,034,992-bp draft genome sequence is G+C rich (63.5%). The genome assembly was annotated with the Rapid Annotation using Subsystems Technology (RAST) server (11). Annotations of protein encoding genes were validated using homology searches on GenBank, eventually improved using Artemis (12). A total of 47 tRNA genes and 4,787 proteins were predicted, corresponding to 85.75% of the assembled sequence. We identified all *Alphaproteobacteria* core genes (13), suggesting that the present sequence is nearly complete.

Of the predicted proteins, 3,523 (73%) were assigned to gene families using OrthoMCL algorithm and database (14) and 2,416 (50.5%) were assigned to biological KEGG pathways using the KAAS server (15). As expected, "*Ca.* Rhodobacter lobularis" harbors the core metabolic pathways such as glycolysis, gluconeogenesis, the tricarboxylic acid cycle (TCA), and the pentose phosphate pathway. The genome also exhibits a tandem cluster of gene transfer agents, the bacteriophage-like elements that mediate horizontal gene transfer found in nearly all *Rhodobacterales* (16).

Comparison with genome sequences available in RAST showed that *Phaeobacter gallaeciensis* (score, 518), *Sagittula stellata* (score, 509), *Roseobacter* sp. AzwK-3b (score, 503), *Rhodobacterales bacterium* Y4I (score, 489), *Roseobacter* sp. SK209-2-6 (score, 482), *Silicibacter pomeroyi* (score, 452), and *Ruegeria pomeroyi* DSS-3 (score, 438), all *Rhodobacterales*, were the closest neighbors of this new species.

Nucleotide sequence accession numbers. The whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. LFTY00000000. The version described in this paper is the first version, LFTY01000000, and consists of the two sequences LFTY01000001 and LFTY01000002.

ACKNOWLEDGMENTS

This work has been carried out thanks to the support of the A*MIDEX project (ANR-11-IDEX-0001-02) funded by the "Investissements d'avenir" French Government program, managed by the French National Research Agency (ANR). This work was supported by the Aix-Marseille University and the Centre national de la recherche scientifique.

We thank the ProfileXpert platform (http://www.profilexpert.fr/) for genome sequencing and the IMBE services of scientific diving and molecular biology for sponge sampling and DNA extraction. We thank the OSU Institute and Spongex project partners for their help, especially Carole

Borchiellini, Jean Vacelet, Alexander Ereskovsky, Emmanuelle Renard, Guillaume Blanc, Laurent Kodjabachian, and Hassiba Belahbib.

REFERENCES

1. **Gloeckner V, Hentschel U, Ereskovsky AV, Schmitt S.** 2013. Unique and species-specific microbial communities in *Oscarella lobularis* and other Mediterranean *Oscarella* species (Porifera: Homoscleromorpha). Mar Biol **160:**781–791. http://dx.doi.org/10.1007/s00227-012-2133-0.
2. **Martin M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17:**10–12. http://dx.doi.org/10.14806/ej.17.1.200.
3. **Peng Y, Leung HC, Yiu SM, Chin FY.** 2012. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28:**1420–1428. http://dx.doi.org/10.1093/bioinformatics/bts174.
4. **Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T.** 2014. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res **24:**1384–1395. http://dx.doi.org/10.1101/gr.170720.113.
5. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol **13:**R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.
6. **Huang X, Madan A.** 1999. CAP3: A DNA sequence assembly program. Genome Res **9:**868–877. http://dx.doi.org/10.1101/gr.9.9.868.
7. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10:**R25. http://dx.doi.org/10.1186/gb-2009-10-3-r25.
8. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
9. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35:**7188–7196. http://dx.doi.org/10.1093/nar/gkm864.
10. **Zhu W, Lomsadze A, Borodovsky M.** 2010. *Ab initio* gene identification in metagenomic sequences. Nucleic Acids Res **38:**e132. http://dx.doi.org/10.1093/nar/gkq275.
11. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.
12. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
13. **Williams KP, Sobral BW, Dickerman AW.** 2007. A robust species tree for the *Alphaproteobacteria*. J Bacteriol **189:**4578–4586. http://dx.doi.org/10.1128/JB.00269-07.
14. **Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ.** 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics **35:**1–19. http://dx.doi.org/10.1002/0471250953.bi0612s35.
15. **Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M.** 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res **35:**W182–W185. http://dx.doi.org/10.1093/nar/gkm321.
16. **Lang AS, Beatty JT.** 2007. Importance of widespread gene transfer agent genes in α-proteobacteria. Trends Microbiol **15:**54–62. http://dx.doi.org/10.1016/j.tim.2006.12.001.