Complete Genome Sequence of *Sphingobium xenophagum* PH3-15, Isolated from La Roche-Posay Thermal Water Sources

Pascal Hilaire,[a] Sandy Contreras,[b] Hélène Blanquart-Goudezeune,[b] Jonathan Verbeke,[c] Baudoin Delépine,[c] Lucas Marmiesse,[c] Rémi Peyraud,[c] Stanislas C. Morand[d]

[a]L'Oréal R&I, Tours, France
[b]GenoScreen, Lille, France
[c]iMean, Toulouse, France
[d]L'Oréal R&I, Aulnay-Sous-Bois, France

ABSTRACT We report here the complete genome sequence of *Sphingobium xenophagum* strain PH3-15, which was isolated from La Roche-Posay thermal water sources. The assembled 4.6-Mbp genome consisted of two chromosomes and three plasmids. These data will provide valuable information and important insights into the physiology and metabolism of this *Sphingobium* organism.

Members of the *Sphingobium* genus are nonpathogenic, aerobic, Gram-negative, lipopolysaccharide-free alphaproteobacteria known to utilize or degrade recalcitrant natural and anthropogenic compounds (1, 2). From the systematics perspective, the genus *Sphingomonas* was renamed *Sphingobium* in 2001 (3), and the species *Sphingomonas xenophaga* (also known by its heterotypic synonym *Sphingomonas hydrophobicum*) was reclassified as *Sphingobium xenophagum* (4, 5). *S. xenophagum* strain PH3-15 was isolated from La Roche-Posay spa water (46°46′46″N, 0°48′27″E) in October 2006; we present here its complete genome sequence.

The strain was isolated on a tryptic soy agar plate and cultured with tryptic soy broth (TSB) medium at 30°C. The initial *Sphingobium* taxonomic identification was deciphered using a BLASTn search on the nonredundant database v2.2.29+ (6) after 16S PCR followed by Sanger sequencing. High-molecular-weight genomic DNA was extracted from an overnight culture grown in TSB at 30°C using the Gentra Puregene kit (Qiagen). Sequencing was first performed on a MinION device using an R9.4.1 flow cell (Oxford Nanopore Technologies; rapid sequencing SQK-RAD004 library with 400 ng DNA and 50-s tagmentation; base calling using Guppy v4.2.2 in 450bps_fast configuration). The long reads (2.16 Gb; 225,196 reads; N_{50}, 19,173 bp; quality score, 9.6) were quality controlled using MinIONQC v1.4.1 (7) and filtered and trimmed using NanoFilt v2.6.0 (-q 7 -l 1000); adaptors were removed using Porechop v0.2.4 (8, 9). Then, sequencing on a HiSeq 4000 system (Illumina; Nextera XT library) generated 150-bp paired-end short reads, which were assessed using FastQC v0.11.5 (10) and cleaned using Cutadapt v1.18 (11) and Prinseq v0.20.4 (12) (parameters: -trim_qual_right 30 -trim_qual_type min -trim_qual_rule lt -trim_qual_window 7 -ns_max_n 0 -noniupac -min_qual_mean 30 -trim_left 15 -min_len 60), resulting in $2 \times 12,235,539$ reads (3,117 Mb). Both short and long cleaned reads were *de novo* assembled using the SPAdes v3.10.1 (13) and MaSuRCA v3.3.0 (PE = pe 600 50) (14) tools. The consensus assembly was manually curated and carefully verified using Bowtie2 v2.1.0 (15), Minimap2 v2.17 (16), and Geneious Assembler (Biomatters) by calculating the reads mapped back to contigs (RMBC) index (98.6 to 99.2% rate for the three tools). After assembly polishing using Pilon v1.23 (17), the chromosomal and plasmid coverage and circularity were validated using Geneious Prime v2020.2.5 software by visualizing the reads that overlapped both the 5′ and 3′ replicon extremities. The whole-genome median coverage depth was 1,272-fold. The assembled ungapped genome is 4,577,807 bp long (GC content, 63.0%), comprising two chromosomes (3,472,664 and 623,015 bp) and three plasmids named

pSH1 (263,324 bp), pSH2 (162,057 bp), and pSH3 (56,747 bp). Structural and functional annotations, carried out using the NCBI Prokaryotic Genome Annotation Pipeline v5.2 (18), identified 3 rRNA operons (5S, 16S, and 23S), 54 tRNAs, and 4,178 protein-coding sequences. Comparison with all other available *Sphingobium* genomes using FastANI v1.31 (19) showed that the closest strain is *Sphingobium hydrophobicum* C1 (average nucleotide identity [ANI], 0.994). Searches for mobile elements using ISsaga v2.0 (20) revealed the presence of 36 *istA* to *istB* operons belonging to the transposase-cointegrase IS*21* family (21). Finally, genes related to exopolysaccharide polymerization and export as well as to nostoxanthin pigment biosynthesis were also recovered (22–24).

Data availability. The complete genome sequence is available in GenBank under the accession numbers CP076556.1 to CP076560.1, BioProject accession number PRJNA600023, and BioSample accession number SAMN13781057. The version described in this paper is the first version. The raw reads have been deposited at the Sequence Read Archive under the accession number SRS8817059.

REFERENCES

1. Stolz A. 2009. Molecular characteristics of xenobiotic-degrading sphingomonads. Appl Microbiol Biotechnol 81:793–811. https://doi.org/10.1007/s00253-008-1752-3.
2. Leys NMEJ, Ryngaert A, Bastiaens L, Verstraete W, Top EM, Springael D. 2004. Occurrence and phylogenetic diversity of Sphingomonas strains in soils contaminated with polycyclic aromatic hydrocarbons. Appl Environ Microbiol 70:1944–1955. https://doi.org/10.1128/AEM.70.4.1944-1955.2004.
3. Takeuchi M, Hamana K, Hiraishi A. 2001. Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. Int J Syst Evol Microbiol 51:1405–1417. https://doi.org/10.1099/00207713-51-4-1405.
4. Chen X, Wang H, Xu J, Song D, Sun G, Xu M. 2016. *Sphingobium hydrophobicum* sp. nov., a hydrophobic bacterium isolated from electronic-waste-contaminated sediment. Int J Syst Evol Microbiol 66:3912–3916. https://doi.org/10.1099/ijsem.0.001287.
5. Feng G-D, Chen M-B, Zhang X-J, Wang D-D, Zhu H-H. 2019. Whole genome sequences reveal the presence of 11 heterotypic synonyms in the genus *Sphingobium* and emended descriptions of *Sphingobium indicum*, *Sphingobium fuliginis*, *Sphingobium xenophagum* and *Sphingobium cupriresistens*. Int J Syst Evol Microbiol 69:2161–2165. https://doi.org/10.1099/ijsem.0.003432.
6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
7. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2018. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35:523–525. https://doi.org/10.1093/bioinformatics/bty654.
8. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
9. Wick RR. 2018 Porechop: adapter trimmer for Oxford Nanopore reads. https://github.com/rrwick/Porechop.
10. Andrews S. 2010 FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
11. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
12. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.
13. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
14. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677. https://doi.org/10.1093/bioinformatics/btt476.
15. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
16. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
17. 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.
18. 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.
19. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.
20. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of Web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol 12:R30. https://doi.org/10.1186/gb-2011-12-3-r30.
21. Berger B, Haas D. 2001. Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS21 and related elements. Cell Mol Life Sci 58:403–419. https://doi.org/10.1007/PL00000866.
22. Wu M, Huang H, Li G, Ren Y, Shi Z, Li X, Dai X, Gao G, Ren M, Ma T. 2017. The evolutionary life cycle of the polysaccharide biosynthetic gene cluster based on the *Sphingomonadaceae*. Sci Rep 7:46484. https://doi.org/10.1038/srep46484.
23. Li H, Zhu H, Sun S, Feng Z, Sun Y, Zhou W. 2016. Biological production of welan gum. Phys Sci Rev 1:e20160049. https://doi.org/10.1515/psr-2016-0049.
24. Zhu L, Wu X, Li O, Qian C, Gao H. 2012. Cloning and characterization of genes involved in nostoxanthin biosynthesis of Sphingomonas elodea ATCC 31461. PLoS One 7:e35099. https://doi.org/10.1371/journal.pone.0035099.