AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

The Genome Sequence of the *Halobacterium salinarum* Type Strain Is Closely Related to That of Laboratory Strains NRC-1 and R1

Friedhelm Pfeiffer,[a] Anita Marchfelder,[b] Bianca Habermann,[a,c] Mike L. Dyall-Smith[a,d]

[a]Computational Biology Group, Max Planck Institute of Biochemistry, Martinsried, Germany
[b]Biology II, Ulm University, Ulm, Germany
[c]CNRS, IBDM UMR 7288, Computational Biology Group, Aix Marseille Université, Marseille, France
[d]Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Australia

ABSTRACT High-coverage long-read sequencing of the *Halobacterium salinarum* type strain (91-R6) revealed a 2.17-Mb chromosome and two large plasmids (148 and 102 kb). Population heterogeneity and long repeats were observed. Strain 91-R6 and laboratory strain R1 showed 99.63% sequence identity in common chromosomal regions and only 38 strain-specific segments. This information resolves the previously uncertain relationship between type and laboratory strains.

Halobacterium salinarum is a well-studied model haloarchaeon first isolated from cured cod in 1922 (1). The source of this organism was found to be salt. This original type strain was lost (2), and a neotype was assigned, *H. salinarum* isolate 91-R6 (3) (NRC 34002 = ATCC 33171 = DSM 3754), which is referred to hereafter as strain 91-R6. It was isolated in Canada from the red discoloration on a salted cow hide (3). Little experimental work has been reported for the type strain, but two laboratory strains of *H. salinarum* have been sequenced, namely, strain NRC-1 (4) and strain R1 (5). Their relationship to the type strain has previously remained uncertain.

A fresh culture of strain 91-R6 (DSM 3754^T) was obtained from the DSMZ and inoculated directly into liquid medium, omitting any colony purification. DNA from the resulting cells was used for genome sequencing using high-coverage PacBio long-read technology (5 single-molecule real-time [SMRT] cells; 253,044 reads; average length, 5,400 bp; and 1.3 Gbp total, using kits from PacBio, including template preparation, MagBead loading, and sequencing). For assembly, we used the SMRTanalysis pipeline (RS_HGAP_assembly.2 v2.3.0, Pacific Biosciences, with default parameters) which runs HGAP (DAGCON-based hierarchical genome assembly process) in three steps (6), namely, preassembly, *de novo* assembly with the Celera assembler, and final polishing with Quiver. Despite high coverage, the assembly gave 43 contigs. A supervised genome assembly was performed using Canu v1.7 (7) for assembly and Geneious v10.2 (8) for integration and editing of contigs. Considerable population heterogeneity (transposon integrations and transposon-triggered genome inversions) was encountered, which explained the failure of the automated assembly procedure. The representative genome sequence consists of 1 chromosome (2,178,608 bp, 67.1% G+C content, and 400-fold coverage) and 2 large plasmids (pHSAL1, 148,406 bp and 60.6% G+C content; pHSAL2, 102,666 bp and 56.5% G+C content; 500-fold coverage for plasmids). The plasmids share a 39,230-bp duplication devoid of any sequence difference.

The chromosomes of strains 91-R6 and R1 were compared in detail by methods we previously described for strain comparisons of *Haloquadratum walsbyi* and *Photorhabdus laumondii* (9, 10). They showed very high DNA sequence similarity (99.63% se-

Citation Pfeiffer F, Marchfelder A, Habermann B, Dyall-Smith ML. 2019. The genome sequence of the *Halobacterium salinarum* type strain is closely related to that of laboratory strains NRC-1 and R1. Microbiol Resour Announc 8:e00429-19. https://doi.org/10.1128/MRA.00429-19.

Editor Kenneth M. Stedman, Portland State University

Address correspondence to Friedhelm Pfeiffer, fpf@biochem.mpg.de.

Received 22 April 2019
Accepted 18 June 2019
Published 11 July 2019

quence identity covering 84.9% of 2.17 Mb in strain 91-R6 and 92.5% of 2.00 Mb in strain R1) and complete colinearity. Only 38 strain-specific regions were identified. As the chromosomes of strains R1 and NRC-1 are nearly identical, there is also very high similarity of the chromosomes from strains 91-R6 and NRC-1. The plasmids of strains 91-R6 and R1 exhibited patches with very high interstrain similarity (107 kb, pHSAL1/pHS3; 42.5 kb, chromosome/pHS3; and 13.3 kb, pHSAL2/pHS1).

Given the close genomic similarity of the strains, the annotation of strain R1 was used as a reference for that of strain 91-R6. Strain 91-R6 codes for 2,451 regular proteins, of which 2,092 are shared with strain R1, with only a minority (73) having less than 98% protein sequence identity. Strain R1 is 1 of 12 haloarchaeal genomes which have been reliably annotated by our gold standard protein-based strategy (5, 11). Our efforts also include regular systematic correlation with high-level databases (Swiss-Prot and KEGG).

During this project, we revised the annotation of the *H. salinarum* NRC-1 genome and submitted it to NCBI as a third-party annotation (NCBI:TPA).

Data availability. The genome sequence of strain 91-R6 has been deposited in GenBank under the accession numbers CP038631 (chromosome), CP038632 (pHSAL1), and CP038633 (pHSAL2). Raw reads have been deposited in the SRA archive under BioProject accession number PRJNA530823. The third-party annotation of the NRC-1 genome has been deposited in the Third Party Annotation section of GenBank (accession numbers BK010829, BK010830, and BK010831).

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Harrison FC, Kennedy ME. 1922. The red discolouration of cured codfish. Proc Trans R Soc Can 16:101–152.
2. Grant WD, Kamekura M, McGenity TJ, Ventosa A. 2001. Class III. *Halobacteria* class. nov., p 294–334. *In* Boone D, Castenholz R, Garrity G (ed), Bergey's manual of systematic bacteriology, 2nd ed. Springer-Verlag, New York, NY.
3. Lochhead AG. 1934. Bacteriological studies on the red discoloration of salted hides. Can J Res 10:275–286. https://doi.org/10.1139/cjr34-026.
4. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ, Hough DW, Maddocks DG, Jablonski PE, Krebs MP, Angevine CM, Dale H, Isenbarger TA, Peck RF, Pohlschroder M, Spudich JL, Jung KW, Alam M, Freitas T, Hou S, Daniels CJ, Dennis PP, Omer AD, Ebhardt H, Lowe TM, Liang P, Riley M, Hood L, DasSarma S. 2000. Genome sequence of Halobacterium species NRC-1. Proc Natl Acad Sci U S A 97:12176–12181. https://doi.org/10.1073/pnas.190337797.
5. Pfeiffer F, Schuster SC, Broicher A, Falb M, Palm P, Rodewald K, Ruepp A, Soppa J, Tittor J, Oesterhelt D. 2008. Evolution in the laboratory: the genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. Genomics 91:335–346. https://doi.org/10.1016/j.ygeno.2008.01.001.
6. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
7. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
8. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
9. Dyall-Smith ML, Pfeiffer F, Klee K, Palm P, Gross K, Schuster SC, Rampp M, Oesterhelt D. 2011. *Haloquadratum walsbyi*: limited diversity in a global pond. PLoS One 6:e20968. https://doi.org/10.1371/journal.pone.0020968.
10. Zamora-Lagos MA, Eckstein S, Langer A, Gazanis A, Pfeiffer F, Habermann B, Heermann R. 2018. Phenotypic and genomic comparison of Photorhabdus luminescens subsp. laumondii TT01 and a widely used rifampicin-resistant Photorhabdus luminescens laboratory strain. BMC Genomics 19:854. https://doi.org/10.1186/s12864-018-5121-z.
11. Pfeiffer F, Oesterhelt D. 2015. A manual curation strategy to improve genome annotation: application to a set of haloarchael genomes. Life (Basel) 5:1427–1444. https://doi.org/10.3390/life5021427.