

UC San Diego

UC San Diego Previously Published Works

Title

Metagenome-assembled genome of withering syndrome causative agent, *Candidatus Xenohaliotis californiensis*, from endangered white abalone (*Haliotis sorenseni*).

Permalink

<https://escholarship.org/uc/item/8m41f4f9>

Journal

Microbiology Resource Announcements, 13(6)

Authors

Kunselman, Emily

Allard, Sarah

Burge, Colleen

et al.

Publication Date

2024-06-11

DOI

10.1128/mra.00086-24

Peer reviewed

Metagenome-assembled genome of withering syndrome causative agent, “*Candidatus Xenohalictis californiensis*,” from endangered white abalone (*Haliotis sorenseni*)

Emily Kunselman,¹ Sarah Allard,^{1,2} Colleen Burge,^{3,4} Blythe Marshman,³ Alyssa Frederick,⁴ Jack Gilbert^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT The genome of “*Candidatus Xenohalictis californiensis*” was assembled from shotgun metagenomic sequencing of experimentally infected white abalone. Ninety-one percent genome completeness was achieved with low contamination. Sequencing this genome provides the opportunity to track pathogen evolution over time, conduct gene expression experiments, and study dynamics between this pathogen and its phage.

KEYWORDS abalone, pathogen, genome

White abalone (*Haliotis sorenseni*) from the Captive Breeding Program at UC Davis Bodega Marine Lab were experimentally infected with the intracellular bacterial pathogen, “*Candidatus Xenohalictis californiensis*” (*CaXc*), using header tanks with infected abalone. *CaXc* is uncultured but can be transferred to other abalone through feces. *CaXc* presence and quantification were determined using *CaXc*-specific quantitative PCR (1) and 16S rRNA amplicon sequencing of the V4 region using primer pair 515F–806R (2). Post-esophagus, digestive gland tissue, and fecal matter were dissected, with some containing over 99% of *Candidatus Xenohalictis* reads. One post-esophagus and one digestive gland sample from different abalone were chosen for shotgun metagenomic sequencing due to the high proportion of pathogen DNA. A fecal sample with 20% of *Candidatus Xenohalictis* reads was also sequenced due to an expectation of lower host DNA contamination. The goal of shotgun metagenomic sequencing was to obtain a genome for the abalone pathogen, *CaXc*, which causes withering syndrome in many abalone species.

DNA was extracted by the UCSD Microbiome Core using the Applied Biosystems MagMax Ultra Nucleic Acid Isolation Kit (cat #A52358). Library preparation was conducted by the UCSD Microbiome Core using the KAPA Hyper Plus Kit (Roche Diagnostics, USA). Sequencing by synthesis was conducted by the UCSD IGM Genomics Center on the Illumina NovaSeq 6000 platform with paired-end 150 base pair cycles.

A protocol for assembling draft genomes from metagenomic sequencing reads was followed in KBase (3). Default parameters were used for all programs. Read quality was assessed with FastQC v0.12.1. A total of 18,524,162 sequences were obtained across all three samples, and sequence quality remained above 30 for all bases. Trimmomatic v0.36 was used to trim adapters and pair forward and reverse reads (4). Following this, 12,667,415 paired sequences were retained. metaSPAdes v3.15.3 was chosen for assembly based on best contig length, high N50, and low L50 (5). Contigs were binned using MaxBin2 v2.2.4, MetaBAT2 v1.7, and CONCOCT v1.1, and DAS tool v1.1.2 was used to optimize bins from all outputs combined (6–9). Three bins were generated, one of which was from the order Rickettsiales and was suspected to be the *CaXc* genome. Using QUAST (10, 11), this bin was found to have 35 contigs and N50 of 64,787.

Editor Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Emily Kunselman, ekunselm@ucsd.edu.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 1 February 2024

Accepted 8 April 2024

Published 30 April 2024

Copyright © 2024 Kunselman et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

To confirm the identity of the assembly as *CaXc*, the 16S rRNA gene sequence was extracted using ConTest16S (12) and was compared to partial *CaXc* sequences using NCBI's BLASTn. The sequence shared 99% similarity across 99% of the query with "abalone withering syndrome agent" and 100% similarity across 88% of the query with "*Candidatus Xenohalictis californiensis*." This confirmed that the genome assembly is that of *CaXc*, the abalone pathogen in our experiment.

N's in the *CaXc* assembly were removed before it was uploaded to MicroScope Microbial Genome Annotation and Analysis Platform v3.16.2 (13). The genome is 1,091,660 base-pairs long with 1,040 protein-coding sequences. CheckM (14) determined that the assembly had 91.1369% completeness and 1.18% contamination with 25 marker genes missing and three markers duplicated. *CaXc* is a gram-negative bacterium, and the genome has a guanine-cytosine content (GC) content of 30.99%.

ACKNOWLEDGMENTS

This work was funded in part through contract P1970003 from CDFW to UC Davis, through a federal NOAA Section 6 Grant NA19NMF4720103 to CDFW. All work with white abalone was performed under the Endangered Species Act (ESA) permit for white abalone (Invertebrate Enhancement Permit 14344-2R) held by UC Davis.

This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (#S10 OD026929).

The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the MicroScope annotation platform.

AUTHOR AFFILIATIONS

¹Scripps Institution of Oceanography, Center for Marine Biotechnology and Biomedicine, La Jolla, California, USA

²Department of Pediatrics, University of California, La Jolla, California, USA

³California Department of Fish and Wildlife, Bodega Bay, California, USA

⁴Bodega Marine Laboratory, University of California Davis, Bodega Bay, California, USA

AUTHOR ORCIDs

Emily Kunselman  <http://orcid.org/0000-0002-3866-5644>

FUNDING

Funder	Grant(s)	Author(s)
DOC National Oceanic and Atmospheric Administration (NOAA)	NA19NMF4720103	Alyssa Frederick

DATA AVAILABILITY

The *CaXc* genome is publicly available through ENA project accession [PRJEB68339](#).

REFERENCES

- Friedman C, Wight N, Crosson L, White S, Strenge R. 2014. Validation of a quantitative PCR assay for detection and quantification of '*Candidatus Xenohalictis californiensis* Dis Aquat Org 108:251–259. <https://doi.org/10.3354/dao02720>
- Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* 75:129–137. <https://doi.org/10.3354/ame01753>
- Chivian D, Jungbluth SP, Dehal PS, Wood-Charlson EM, Canon RS, Allen BH, Clark MM, Gu T, Land ML, Price GA, Riehl WJ, Sneddon MW, Sutormin R, Zhang Q, Cottingham RW, Henry CS, Arkin AP. 2023. Metagenome-assembled genome extraction and analysis from microbiomes using KBase. *Nat Protoc* 18:208–238. <https://doi.org/10.1038/s41596-022-00747-x>

4. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
5. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>
6. Wu Y-W, Simmons BA, Singer SW. 2016. Maxbin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>
7. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. Metabat 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
8. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>
9. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
10. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
11. Mikheenko A, Valin G, Pribelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for *de novo* assembly evaluation. *Bioinformatics* 32:3321–3323. <https://doi.org/10.1093/bioinformatics/btw379>
12. Lee I, Chalita M, Ha S-M, Na S-I, Yoon S-H, Chun J. 2017. Contest16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int J Syst Evol Microbiol* 67:2053–2057. <https://doi.org/10.1099/ijsem.0.001872>
13. Vallenet D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, Burlot L, Bussell X, Fouteau S, Gautreau G, Lajus A, Langlois J, Planel R, Roche D, Rollin J, Rouy Z, Sabatet V, Médigue C. 2019. Microscope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res*. gkz926. <https://doi.org/10.1093/nar/gkz926>
14. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>