1 # Closing the genome of *Teredinibacter turnerae* T7902 by long-read
2 nanopore sequencing
3

4 Mark T. Gasser[a] | Annie Liu[a] | Ron Flatau[b] | Marvin Altamia[b] | Claire Marie Filone[a] | Dan

5 Distel[b#]

6

7 [a]Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA 20723

8 [b]Ocean Genome Legacy Center, Northeastern University, Nahant, Massachusetts, USA 01908

9

10 Running Title: Closing the genome of *T. turnerae* T7902

11

12 MTG: mark.gasser@jhuapl.edu (ORCID: 0000-0002-9396-4954)

13 AL: annie.liu@jhuapl.edu (ORCID: 0009-0002-3224-6008)

14 RF: r.flatau@northeastern.edu (ORCID: 0000-0002-4357-5870)

15 MAA: marvin.altamia@gmail.com (ORCID: 0000-0002-8625-767X)

16 CMF: claire.marie.filone@jhuapl.edu (ORCID: 0000-0002-4041-3948)

17 DLD: d.distel@northeastern.edu (ORCID: 0000-0002-3860-194X) #Corresponding author

18

19 ## Abstract
20 We present the complete closed circular genome sequence derived from Oxford Nanopore

21 sequencing of the shipworm endosymbiont *Teredinibacter turnerae* T7902 (DSM 15152, ATCC

22 39867), originally isolated from the shipworm *Lyrodus pedicellatus* (1). This sequence will aid in

23   the comparative genomics of shipworm endosymbionts and the understanding of host-

24   symbiont evolution.

25   Announcement

26   *Teredinibacter* species are cellulolytic gammaproteobacteria (Cellvibrionaceae) that occur as

27   intracellular endosymbionts of wood-boring bivalves (Teredinidae) (1–4), commonly known as

28   shipworms. Strain T7902 was isolated from the gills of a single specimen of the shipworm

29   *Lyrodus pedicellatus* collected in Long Beach, CA, in 1979 and was the second strain of *T.*

30   *turnerae* brought into pure culture (1). It is the original representative of *T. turnerae* Clade II,

31   one of two distinct clades previously identified among *T. turnerae* strains (5, 6). Gills were

32   dissected, washed in sterile seawater, and homogenized in shipworm basal medium (SBM) (7),

33   and the homogenate was streaked onto a culture plate containing 0.9% agar and SBM at pH 8.0

34   supplemented with 0.2% w/v powdered cellulose (Sigmacell Type 100; Sigma-Aldrich) and 5

35   mM ammonium chloride. Individual colonies were picked and restreaked on fresh plates until a

36   pure clonal isolate was obtained. The original genome sequence of *T.turnerae* T7902 was

37   published to GenBank (GCA_000379165.1) but has not been described previously in peer-

38   reviewed literature. This sequence was completed on 2012-05-22 at the DOE Joint Genome

39   Institute under award 10.46936/10.25585/60001419 using 454 GS FLX Titanium and Illumina

40   HiSeq 2000 sequencing platforms. It was assembled using Velvet v. 1.0.13 (8) and ALLPATHS v.

41   R40295 (9), resulting in an improved high-quality draft assembly comprised of 72 scaffolds with

42   76 contigs. At the time of this work, the genomes of nine strains of *T. turnerae* were publicly

43   available at the National Center for Biotechnology Information (NCBI), and only one strain

44   genome, T7901 (Clade I, Genbank: GCA_000023025.1), was complete and closed. Here, we

45   present the re-sequencing and completed genome of strain T7902 from nanopore-only

46   sequencing (Table 1).

47   *Table 1 Assemblies of Teredinibacter turnerae T7902*

GenBank Assembly	Scaffolds (contigs)	Size (bp)	GC%	CDS
GCA_000379165.1	72 (76)	5,387,817	50.8	4268
GCA_037935975.1 (this work)	1 (1)	5,348,823	50.9	4212

48

49   *Teredinibacter turnerae* strain T7902 colonies were maintained at 30°C on shipworm basal

50   medium (SBM) (7) plates supplemented with 0.025% NH_4Cl and 0.2% cellulose (Sigmacell Type

51   101; Sigma-Aldrich). A colony was selected and used to inoculate a 6 mL liquid culture of SBM

52   supplemented with 0.025% NH_4Cl and 0.2% carboxymethyl cellulose (CMC) medium and grown

53   at 30°C, 100 rpm for 4 days. Bacterial cells were harvested by centrifugation (10 minutes, 4°C,

54   4,000 × g), and high-molecular-weight DNA was isolated from the cell pellet using the Wizard

55   HMW DNA Extraction Kit (Promega, US) according to the manufacturer's protocol. DNA quality

56   and length were assessed on Tapestation (Agilent Technologies, US). Nanopore (Oxford

57   Nanopore Technologies, UK) sequencing was performed on unsheared gDNA with no size

58   selection and prepared using Nanopore Q20+ chemistry kit v14 and sequenced on a MinION

59   instrument using an R10.4 (FLO-MIN112) flow cell. The standard ligation protocol for the

60   selected kit was followed with the selection of the long fragment buffer (LFB). Bases were

61   called using Guppy v6.5.7 with the super-accurate (SUP) algorithm with default quality read

62   filtering, generating 300,309 reads and an N_{50} read length of 8,763 bp. De novo assembly was

63   performed with Flye v2.9.2 (https://github.com/fenderglass/Flye) (10) followed by contig

64   correction and consensus generation with Medaka v1.8.0

65   (https://github.com/nanoporetech/medaka). To generate a circular chromosome, overlaps

66   were identified and removed before the assembly was rotated to the gene predicted by

67   prodigal v2.6.3 (11) nearest the middle of the contig with Circlator v1.5.5

68   (https://github.com/sanger-pathogens/circlator) (12). A closed, circular chromosomal assembly

69   with a genome coverage of 278.0x was produced and annotated using the NCBI Prokaryotic

70   Genome Annotation Pipeline (PGAP) (13). The new assembly is 99.99% (ANI) (14) identical in

71   primary sequence and highly syntenic (15) with the original (Fig. 1). For example, the new

72   assembly and annotation identify 9 ribosomal RNAs, similar to *T. turnerae* strain T7901

73   (Genbank: GCA_000023025.1, ASM2302v1). However, the new assembly reduces the genome

74   size by 38,994 bp to 5,348,823 bp, contains 56 fewer predicted CDS, and resolves several

75   assembly errors. For all software, default parameters were used except where otherwise noted.
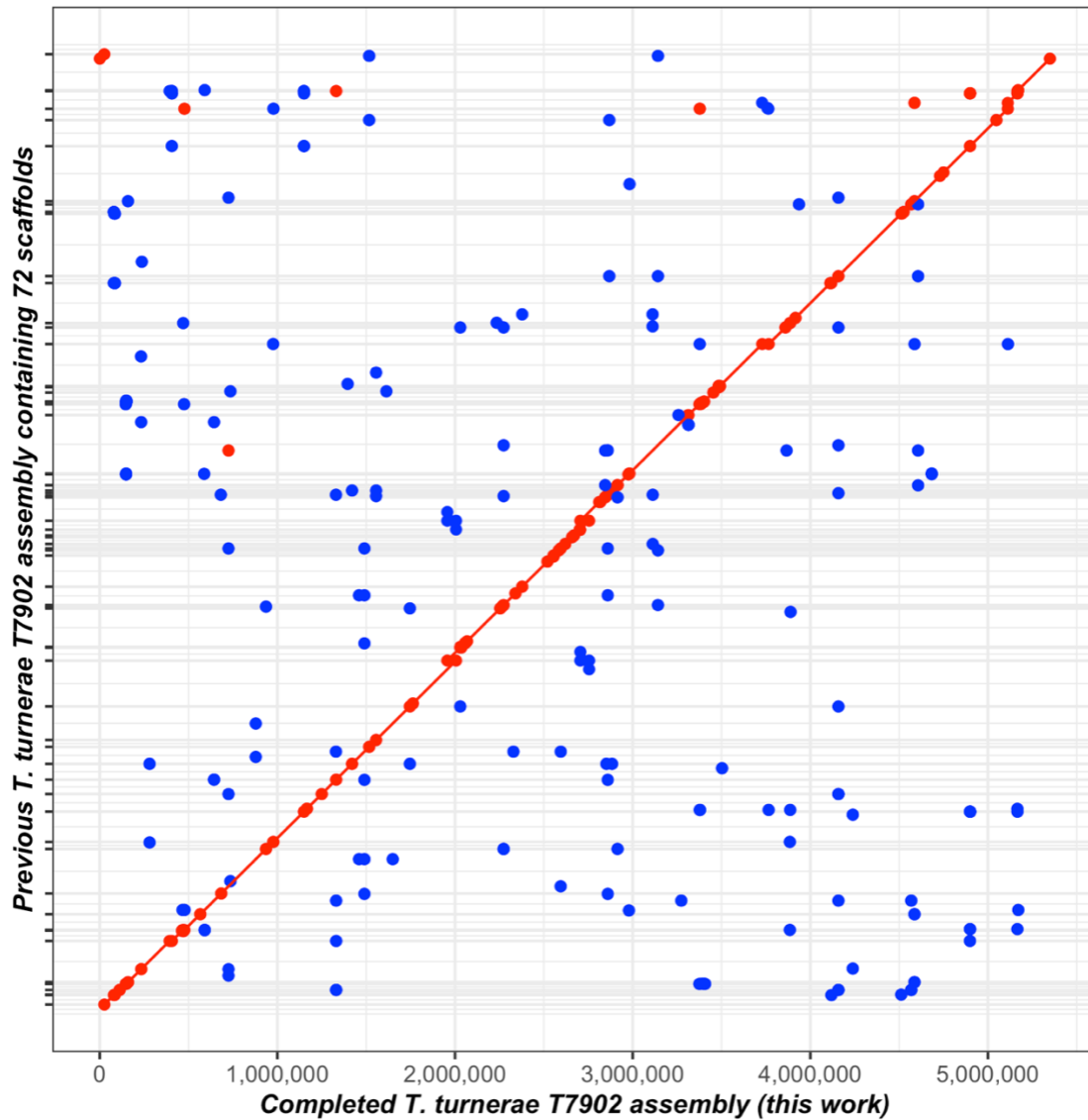
Figure 1. Synteny plot comparing the previously published genome of T. turnerae T7902 (GCA_000379165.1) and the new genome sequence and assembly presented here (GCA_037935975.1). A MUMmer3 plot was generated with NUCmer v3.1 (15) using default settings to assess synteny and completion. Minimum exact matches of 20bp are represented as a dot with lines representing exact match lengths >20bp. Forward matches are displayed in red, while reverse matches are shown in blue.

Data availability

The complete genome sequence of T7902 has been deposited in GenBank under the accession

number CP149817. The Oxford Nanopore sequencing reads are available from the NCBI

Sequence Read Archive (SRA) under the accession number SRR28421272.

94   References

95   1. Distel DL, Morrill W, MacLaren-Toussaint N, Franks D, Waterbury J. 2002. Teredinibacter

96   turnerae gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-

97   proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae).

98   International Journal of Systematic and Evolutionary Microbiology 52:2261–2269.

99   2. Altamia MA, Shipway JR, Stein D, Betcher MA, Fung JM, Jospin G, Eisen J, Haygood MG,

100   Distel DL. 2020. Teredinibacter waterburyi sp. nov., a marine, cellulolytic endosymbiotic

101   bacterium isolated from the gills of the wood-boring mollusc Bankia setacea (Bivalvia:

102   Teredinidae) and emended description of the genus Teredinibacter. International Journal of

103   Systematic and Evolutionary Microbiology 70:2388–2394.

104   3. Altamia MA, Shipway JR, Stein D, Betcher MA, Fung JM, Jospin G, Eisen J, Haygood MG,

105   Distel DL. 2021. Teredinibacter haidensis sp. nov., Teredinibacter purpureus sp. nov. and

106   Teredinibacter franksiae sp. nov., marine, cellulolytic endosymbiotic bacteria isolated from

107   the gills of the wood-boring mollusc Bankia setacea (Bivalvia: Teredinidae) and emended

108   description of the genus Teredinibacter. International Journal of Systematic and

109   Evolutionary Microbiology 71.

110 4. Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedamallu CS, Hostetler JB, Radune D, Toms BS,

111   Henrissat B, Coutinho PM, Schwarz S, Field L, Trindade-Silva AE, Soares CAG, Elshahawi S,

112   Hanora A, Schmidt EW, Haygood MG, Posfai J, Benner J, Madinger C, Nove J, Anton B,

113   Chaudhary K, Foster J, Holman A, Kumar S, Lessard PA, Luyten YA, Slatko B, Wood N, Wu B,

114   Teplitski M, Mougous JD, Ward N, Eisen JA, Badger JH, Distel DL. 2009. The Complete

115   Genome of Teredinibacter turnerae T7901: An Intracellular Endosymbiont of Marine Wood-

116   Boring Bivalves (Shipworms). PLoS ONE 4:e6085.

117 5. Altamia MA, Lin Z, Trindade-Silva AE, Uy ID, Shipway JR, Wilke DV, Concepcion GP, Distel DL,

118   Schmidt EW, Haygood MG. 2020. Secondary Metabolism in the Gill Microbiota of

119   Shipworms (Teredinidae) as Revealed by Comparison of Metagenomes and Nearly

120   Complete Symbiont Genomes. mSystems 5:e00261-20.

121 6. Altamia MA, Wood N, Fung JM, Dedrick S, Linton EW, Concepcion GP, Haygood MG, Distel

122   DL. 2014. Genetic differentiation among isolates of *Teredinibacter turnerae,* a widely

123   occurring intracellular endosymbiont of shipworms. Molecular Ecology 23:1418–1432.

124 7. Waterbury JB, Calloway CB, Turner RD. 1983. A Cellulolytic Nitrogen-Fixing Bacterium

125   Cultured from the Gland of Deshayes in Shipworms (Bivalvia: Teredinidae). Science

126   221:1401–1403.

127 8. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de

128   Bruijn graphs. Genome Res 18:821–829.

129 9. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe

130   DB. 2008. ALLPATHS: De novo assembly of whole-genome shotgun microreads. Genome

131   Res 18:810–820.

132 10. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using

133   repeat graphs. Nat Biotechnol 37:540–546.

134 11. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic

135   gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

136 12. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated

137   circularization of genome assemblies using long sequencing reads. Genome Biol 16:294.

138 13. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F,

139   Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J,

140   Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq:

141   expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model

142   curation. Nucleic Acids Research 49:D1020–D1028.

143 14. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for

144   specialized analyses of microbial genomes and metagenomes. preprint. PeerJ Preprints.

145   15. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004.

146   Versatile and open software for comparing large genomes. Genome Biol 5:R12.

147