AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Chromosome Sequences of a Clinical Isolate of the Free-Living Ameba *Naegleria fowleri*

Ibne Karim M. Ali,[a] Alyssa Kelley,[a,e] Sandeep J. Joseph,[a*] Subin Park,[a,c] Shantanu Roy,[a] Jonathan Jackson,[a,b] Jennifer R. Cope,[a] Lori A. Rowe,[d] Mark Burroughs,[d] Mili Sheth,[d] Dhwani Batra,[d] Vladimir Loparev[d]

[a]Waterborne Disease Prevention Branch, DFWED/NCEZID, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[b]IHRC, Inc., Atlanta, Georgia, USA
[c]Eagle Medical Services, Atlanta, Georgia, USA
[d]Biotechnology Core Facility Branch, DSR/NCEZID, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[e]ASRT, Inc., Smyrna, Georgia, USA

Ibne Karim M. Ali and Alyssa Kelley contributed equally to this work. Author order was determined alphabetically.

ABSTRACT We present the chromosome sequences of a *Naegleria fowleri* isolate from a human primary amebic meningoencephalitis (PAM) case. The genome sequences were assembled from Illumina HiSeq and PacBio sequencing data and verified with the optical mapping data. This led to the identification of 37 contigs representing 37 chromosomes in *N. fowleri*.

The free-living ameba *Naegleria fowleri* (phylum Percoloza; class Heterolobosea) is distributed worldwide in warm freshwater (1, 2). It causes a highly (>97%) fatal brain infection known as primary amebic meningoencephalitis (PAM) (3, 4). The pathogenicity and biology of *N. fowleri* are poorly understood partly due to a lack of good-quality genome sequences. Two currently available genome sequences of *N. fowleri* are fragmented and dispersed in 1,124 contigs (5) and 81 contigs (6), respectively. Here, we identified the chromosome sequences of a reference *N. fowleri* strain (isolate TY).

The *N. fowleri* isolate TY was from a 1969 primary amebic meningoencephalitis (PAM) patient from Virginia and belongs to genotype III (7). It was maintained under axenic culture conditions. For Illumina HiSeq sequencing, DNA was extracted from approximately 2×10^6 amebas with the DNeasy blood and tissue kit (Qiagen, USA). For library preparation, DNA was sheared using a Covaris LE220 focused ultrasonicator (Covaris, Inc., USA) and cleaned with the AMPure kit (Beckman Coulter, Inc., USA). This was followed by dual-indexing using the NEBNext Ultra library prep reagents (New England Biolabs, USA) and barcoding. Libraries were analyzed for size and concentration, pooled, and denatured for loading onto flow cells for cluster generation. Sequencing was performed on an Illumina HiSeq 2500 instrument using HiSeq rapid SBS v2 250 × 250-cycle paired-end sequencing kits. Illumina reads were filtered using BBDuk (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/) to remove adapters, quality trim to a minimum PHRED score of 20, and remove reads shorter than 50 bp. Additionally, high-molecular-weight (HMW) genomic DNA was purified using the MagAttract HMW DNA kit (Qiagen, USA) and sequenced using the PacBio platform (Pacific Biosciences, USA). DNA libraries were prepared using the SMRTbell template prep kit 1.0 and polymerase binding kit P6, and the filtered reads (minLength = 1,000 bp) were *de novo* assembled using Canu v1.6. PacBio's SmrtAnalysis software was used (with default parameters) for adapter trimming and quality control (QC). The resulting consensus sequences were determined with Quiver v1 (https://github.com/PacificBiosciences/GenomicConsensus). The assembly was refined and confirmed by comparison to restriction digest optical maps using the Argus system (OpGen) with MapSolver v2.1.1. The final PacBio assembly of *N. fowleri* TY was further

Citation Ali IKM, Kelley A, Joseph SJ, Park S, Roy S, Jackson J, Cope JR, Rowe LA, Burroughs M, Sheth M, Batra D, Loparev V. 2021. Draft chromosome sequences of a clinical isolate of the free-living ameba *Naegleria fowleri*. Microbiol Resour Announc 10:e01034-20. https://doi.org/10.1128/MRA.01034-20.

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Ibne Karim M. Ali, iali@cdc.gov.

* Present address: Sandeep J. Joseph, Laboratory Reference and Research Branch, Division of Sexually Transmitted Disease Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

Received 6 October 2020
Accepted 14 March 2021
Published 15 April 2021

TABLE 1 Chromosome sizes and telomere lengths in the *N. fowleri* TY isolate

Chromosome name	Chromosome size (bp)	Telomere length (bp)	
		5′ telomere	3′ telomere
Chr01	1,206,962	784	699
Chr02	1,059,554	640	540
Chr03	985,518	724	615
Chr04	984,171	322	518
Chr05	973,969	382	30
Chr06	961,058	657	840
Chr07	853,063	69	377
Chr08	843,297	483	384
Chr09	840,958	625	710
Chr10	833,951	723	618
Chr11	830,534	452	96
Chr12	822,913	477	456
Chr13	800,153	349	770
Chr14	775,976	159	580
Chr15	758,505	687	400
Chr16	756,811	646	734
Chr17	745,360	697	819
Chr18	739,698	664	281
Chr19	725,748	770	449
Chr20	717,112	590	665
Chr21	716,195	581	599
Chr22	713,437	727	861
Chr23	711,928	545	522
Chr24	704,950	127	530
Chr25	667,433	702	703
Chr26	665,971	440	758
Chr27	647,405	595	682
Chr28	638,607	538	498
Chr29	626,360	261	709
Chr30	626,308	610	0
Chr31	614,393	16	162
Chr32	601,738	633	623
Chr33	596,164	248	394
Chr34	592,949	689	512
Chr35	570,195	1,022	451
Chr36	547,731	514	584
Chr37	537,351	360	383

polished by mapping Illumina reads using the Unicycler polish function (default parameters) in Unicycler v4.4 (8). Default software parameters were used except where otherwise noted.

Mapping the HiSeq Illumina reads (total reads, 5,354,858) to the final PacBio assembly (N_{50}, 10,674; total reads, 396,362) covered 99.97% of the genome with an average coverage depth of 31×. The final assembly of the nuclear genome consisted of 37 contigs comprising 27,994,426 bp, with an N_{50} value of 756,811 bp and GC content of 36.85%. Each contig contained the characteristic N-terminal and C-terminal telomere sequences, suggesting that these were full-length chromosomes except for one contig for which the C-terminal telomere could not be identified (Table 1). The largest contig/chromosome was 1,206,962 bp, and the shortest was 537,351 bp. A total of 9,405 protein-encoding genes were predicted using BRAKER2 (9). Approximately 75% of the genes contain one or more introns. The average length of an intron is 186 bp. Approximately 5.28% of the total genome contains repetitive sequences as detected using RepeatMasker v4.0.8 and RepeatModeler (default parameters). Discovery of chromosome sequences will help scientists better understand the biology and pathogenicity of this ameba. They will help researchers identify virulence factors in *N. fowleri* and effective drug targets for treating PAM patients.

Data availability. The *Naegleria fowleri* TY contig/chromosome sequences have been deposited at the National Center for Biotechnology Information's Sequence Read Archive (SRA) under the accession numbers CP062075, CP062076, CP062077, CP062078, CP062079, CP062080, CP062081, CP062082, CP062083, CP062084, CP062085, CP062086, CP062087, CP062088, CP062089, CP062090, CP062091, CP062092, CP062093, CP062094, CP062095, CP062096, CP062097, CP062098, CP062099, CP062100, CP062101, CP062102, CP062103, CP062104, CP062105, CP062106, CP062107, CP062108, CP062109, CP062110, and CP062111. The raw data are available under the BioSample accession number SRS7483196.

ACKNOWLEDGMENTS

This work was partially supported by an Incubator AMD fund (to Ibne Karim M. Ali), and the efforts of Ibne Karim M. Ali, Sandeep J. Joseph, Shantanu Roy, Jennifer R. Cope, Lori A. Rowe, Mark Burroughs, Mili Sheth, Dhwani Batra, and Vladimir Loparev were funded by the HHS/Centers for Disease Control and Prevention (CDC).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention.

REFERENCES

1. Martinez AJ. 1993. Free-living amebas: infection of the central nervous system. Mt Sinai J Med 60:271–278.
2. Martinez AJ, Visvesvara GS. 1997. Free-living, amphizoic and opportunistic amebas. Brain Pathol 7:583–598. https://doi.org/10.1111/j.1750-3639.1997.tb01076.x.
3. Carter RF. 1970. Description of a *Naegleria* sp. isolated from two cases of primary amoebic meningo-encephalitis, and of the experimental pathological changes induced by it. J Pathol 100:217–244. https://doi.org/10.1002/path.1711000402.
4. Yoder JS, Eddy BA, Visvesvara GS, Capewell L, Beach MJ. 2010. The epidemiology of primary amoebic meningoencephalitis in the USA, 1962–2008. Epidemiol Infect 138:968–975. https://doi.org/10.1017/S0950268809991014.
5. Zysset-Burri DC, Müller N, Beuret C, Heller M, Schürch N, Gottstein B, Wittwer M. 2014. Genome-wide identification of pathogenicity factors of the free-living amoeba *Naegleria fowleri*. BMC Genomics 15:496. https://doi.org/10.1186/1471-2164-15-496.
6. Liechti N, Schürch N, Bruggmann R, Wittwer M. 2019. Nanopore sequencing improves the draft genome of the human pathogenic amoeba *Naegleria fowleri*. Sci Rep 9:16040. https://doi.org/10.1038/s41598-019-52572-0.
7. Zhou L, Sriram R, Visvesvara GS, Xiao L. 2003. Genetic variations in the internal transcribed spacer and mitochondrial small subunit rRNA gene of *Naegleria* spp. J Eukaryot Microbiol 50 Suppl:522–526. https://doi.org/10.1111/j.1550-7408.2003.tb00617.x.
8. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
9. Katharina JH, Simone L, Alexandre L, Mark B, Mario S. 2016. BRAKER1: unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661.