AMERICAN SOCIETY FOR MICROBIOLOGY

Microbiology® Resource Announcements

Check for updates

# Complete Genome Sequence of the *Naegleria fowleri* (Strain LEE) Closed Circular Extrachromosomal Ribosomal DNA Element

John C. Mullican,[a]* Kristen M. Drescher,[b] Nora M. Chapman,[a] Steven Tracy[a]

[a]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, USA
[b]Department of Medical Microbiology and Immunology, Creighton University Medical School, Omaha, Nebraska, USA

**ABSTRACT** The circular extrachromosomal ribosomal DNA (rDNA) element of *Naegleria fowleri* strain LEE was molecularly cloned and fully sequenced. The element comprises 15,786 bp and contains a single copy of the organism's rDNA cistron. The nonribosomal sequence contains five potential open reading frames, two large direct repeat sequences, and numerous smaller repeated-sequence regions.

*N*aegleria fowleri (Heterolobosea, Schizopyrenida, Vahlkampfidae, Naegleria) causes primary amebic meningoencephalitis, a usually fatal brain infection (1–3). *Naegleria* species contain a single ribosomal DNA (rDNA) cistron (5.8S, 18S, and 28S rRNA genes) on closed circular extrachromosomal rDNA elements (CEREs); there are approximately 4,000 copies per cell (4–7). Genome sequencing of three *Naegleria* species (8–11) confirmed that there are no rDNA genes in the nuclear genome. Here, we report the complete sequence of the *N. fowleri* (strain LEE) CERE. While the total genome of *N. fowleri* (strain LEE) has also been sequenced (10), this report demarcates for the first time the complete rRNA coding sequences of *N. fowleri* and identifies five putative open reading frames (ORFs) and repeated sequences in the nonribosomal sequence (NRS).

*N. fowleri* LEE strain amebae, kindly provided by David John (Oklahoma State University, Tulsa, OK, USA), were propagated axenically at 37°C in stationary tissue culture flasks in Nelson's medium (12) supplemented with 4% (by volume) calf serum. Total DNA was extracted from log-phase *N. fowleri* trophozoites, collected centrifugally, and suspended and lysed in a highly chaotropic lysis buffer (8 M NaSCN, 50 mM Tris-HCl [pH 7.5], 50 mM EDTA [pH 8.0], 5 mM EGTA, and 142 mM $\beta$-mercaptoethanol) as described previously (7). Following electrophoresis in 0.8% agarose gels in 0.2× Tris-acetate-EDTA (TAE) buffer and staining with ethidium bromide, the 16-kbp supercoiled CERE, migrating in front of the genomic DNA band, was visualized by UV illumination, excised, and purified free from agarose. The isolated CERE was linearized with HindIII and ligated into the pGEM7Zf(+) vector (Promega Corp.).

The complete sequence was assembled from clones, subclones, and purified uncloned DNA. Illumina next-generation sequencing (NGS) using the NextSeq system was outsourced to Applied Biological Materials (Richmond, BC, Canada), and Sanger sequencing (13) using AB 3500xL genetics analyzers was outsourced to Eurofins Genomics (Louisville, KY). DNA for Illumina sequencing was subjected to tagmentation and adaptor-mediated amplification using the Nextera XT DNA library preparation kit (Illumina), following the vendor's protocol. The final size selection to remove adapter dimers was performed using magnetic beads for DNA purification (Applied Biological Materials). Following library preparation, the samples were run on an Agilent Bioanalyzer using the DNA 1000 kit (Agilent) to confirm the size distribution (a bell curve distribution with a peak between 400 and 600 bp) and to ensure that all adapter dimers had been removed (no peak visible in the 120- to 150-bp range). Samples were quantified by quantitative PCR using the KAPA library quantification kit for the Illumina platform prior to pooling
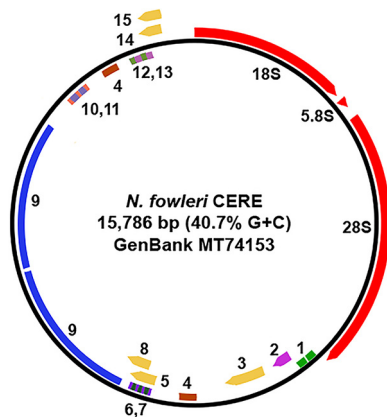
Address correspondence to John C. Mullican, john.mullican@washburn.edu.

* Present address: John C. Mullican, Department of Biology, Washburn University, Topeka, Kansas, USA.

**FIG 1** Map of the CERE from *N. fowleri* strain LEE. Features of the CERE are as listed for GenBank accession number MT741533. 18S, 5.8S, and 28S refer to the rRNA gene coding sequences. Repetitive DNAs are indicated as follows: 1, 168-bp tandem direct repeat; 4, 248-bp direct repeat; 6 and 7, 47- and 56-bp direct repeats, respectively; 9, 2,135-bp direct repeat; 10 and 11, 70- and 130-bp direct repeats, respectively; 12 and 13, 35- and 85-bp direct repeats, respectively. 2 indicates the group 1 intron. ORFs on the sense (or rRNA-encoding) strand are labeled 3, 5, and 8, and ORFs on the antisense strand are labeled 14 and 15.

and sequencing on a NextSeq 500 instrument (Illumina). NGS produced 554,274 paired-end reads of 72-bp average length, totaling 81 Mb; Sanger sequencing produced 118 reads totaling 91.7 kb, representing ~6× coverage in both directions.

The first 25,000 pairs of Illumina reads, representing optimal ~200× coverage of the target plasmid, were assembled using SPAdes version 3.11.0 (removing reads of <33 bases and reads with a Phred score of <20) (14) as implemented in MacVector version 16.0.10 (MacVector, Inc., Cary, NC), resulting in a single linear contig of 17,938 bp. The assembly software utilized default parameters unless otherwise noted. Attempted validation of the consensus using the Align to Reference tool in MacVector with the Illumina reads indicated multiple regions in which the assembly was likely inaccurate due to the presence of multiple direct and inverted repeats. Within MacVector, Sanger reads, which had been base called using Phred, were used in conjunction with the Align to Reference function and the Phrap assembler (version 1.090518) to resolve the repeats and to generate a final circular plasmid sequence of 15,786 bp after removal of the cloning vector sequences.

The 18S, 5.8S, and 28S rRNA gene sequences were identified by search and alignment of Rfam families (15) using the Rfam 14.2 database. ORFs of ≥100 codons on both strands were identified (16). Repeated sequences were identified by BLAST (17) analysis of the entire sequence versus itself using MegaBLAST settings.

An overview of the CERE features, including rRNA genes, ORFs, and repeat sequences, is depicted in Fig. 1. The CERE comprises 15,786 bp (overall GC content, 40.7%). The rDNA region is 5,863 bp (GC content, 46.5%), including two internal transcribed spacers, while the NRS is 9,923 bp (GC content, 37.6%). The NRS contains 10 repetitive DNA sequence families (expect values from 0.0 to 2e−9), accounting for 6,488 bp (65.3%) of the NRS and 41.1% of the total CERE DNA. Repeats range in size from 35 bp to 2,135 bp.

The NRS contains five putative ORFs, three on the rRNA-encoding strand and two on the opposite strand. Interestingly, only one of these (at nucleotides 6809 to 7435) is conserved (84% for DNA and 78% for the predicted amino acid sequence) in the CERE NRS of *Naegleria lovaniensis*, the species most closely related to *N. fowleri* (9), suggesting that this ORF alone may potentially encode a protein.

**Data availability.** The complete assembled sequence has been deposited in GenBank under the accession number MT741533. The version described in this paper represents the first version, MT741533.1. The publicly available raw sequence data for

Sanger and Illumina NGS reads have been deposited in the Sequence Read Archive (SRA) under the accession numbers SRX8926575 and SRX8926574, respectively, and collectively under BioProject accession number PRJNA656555.

## REFERENCES

1. Fowler M, Carter RF. 1965. Acute pyogenic meningitis probably due to *Acanthamoeba* sp.: a preliminary report. Br Med J 2:734–742. https://doi.org/10.1136/bmj.2.5464.734-a.
2. Grace E, Asbill S, Virga K. 2015. *Naegleria fowleri*: pathogenesis, diagnosis, and treatment options. Antimicrob Agents Chemother 59:6677–6681. https://doi.org/10.1128/AAC.01293-15.
3. Piñero JE, Chávez-Munguía B, Omaña-Molina M, Lorenzo-Morales J. 2019. *Naegleria fowleri*. Trends Parasitol 35:848–849. https://doi.org/10.1016/j.pt.2019.06.011.
4. Clark CG, Cross GA. 1987. rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair plasmid. Mol Cell Biol 7:3027–3031. https://doi.org/10.1128/mcb.7.9.3027.
5. Clark CG. 1990. Genome structure and evolution of *Naegleria* and its relatives. J Protozool 37:2S–6S. https://doi.org/10.1111/j.1550-7408.1990.tb01138.x.
6. Clark CG, Cross GA, De Jonckheere JF. 1989. Evaluation of evolutionary divergence in the genus *Naegleria* by analysis of ribosomal DNA plasmid restriction patterns. Mol Biochem Parasitol 34:281–296. https://doi.org/10.1016/0166-6851(89)90057-1.
7. Mullican JC, Chapman NM, Tracy S. 2019. Mapping the single origin of replication in the *Naegleria gruberi* extrachromosomal DNA element. Protist 170:141–152. https://doi.org/10.1016/j.protis.2019.02.001.
8. Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, Shu S, Neupane R, Cipriano M, Mancuso J, Tu H, Salamov A, Lindquist E, Shapiro H, Lucas S, Grigoriev IV, Cande WZ, Fulton C, Rokhsar DS, Dawson SC. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell 140:631–642. https://doi.org/10.1016/j.cell.2010.01.032.
9. Liechti N, Schurch N, Bruggmann R, Wittwer M. 2018. The genome of *Naegleria lovaniensis*, the basis for a comparative approach to unravel pathogenicity factors of the human pathogenic amoeba *N. fowleri*. BMC Genomics 19:654. https://doi.org/10.1186/s12864-018-4994-1.
10. 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.
11. 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.
12. Weik RR, John DT. 1977. Agitated mass cultivation of *Naegleria fowleri*. J Parasitol 63:868–871.
13. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467. https://doi.org/10.1073/pnas.74.12.5463.
14. 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.
15. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, Petrov AI. 2018. Non-coding RNA analysis using the Rfam database. Curr Protoc Bioinformatics 62:e51. https://doi.org/10.1002/cpbi.51.
16. Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102–1104. https://doi.org/10.2144/00286ir01.
17. 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.