

# Draft Genome Sequences from *Cyclospora cayetanensis* Oocysts Purified from a Human Stool Sample

Yvonne Qvarnstrom,<sup>a</sup> Yuping Wei-Pridgeon,<sup>a,b</sup> Wen Li,<sup>a,b</sup> Fernanda S. Nascimento,<sup>a</sup> Henry S. Bishop,<sup>a</sup> Barbara L. Herwaldt,<sup>a</sup> Delynn M. Moss,<sup>c</sup> Vishal Nayak,<sup>c</sup> Ganesh Srinivasamoorthy,<sup>c</sup> Mili Sheth,<sup>c</sup> Michael J. Arrowood<sup>c</sup>

Center for Global Health<sup>a</sup> and National Center for Emerging and Zoonotic Infectious Diseases,<sup>c</sup> Centers for Disease Control and Prevention, Atlanta, Georgia, USA; IHRC, Inc., Atlanta, Georgia, USA<sup>b</sup>

The parasite *Cyclospora cayetanensis* causes foodborne diarrheal illness. Here, we report draft genome sequences obtained from *C. cayetanensis* oocysts purified from a human stool sample. The genome assembly consists of 865 contigs with a total length of 44,563,857 bases. These sequences can facilitate the development of subtyping tools to aid outbreak investigations.

Received 23 September 2015 Accepted 8 October 2015 Published 19 November 2015

**Citation** Qvarnstrom Y, Wei-Pridgeon Y, Li W, Nascimento F, Bishop HS, Herwaldt BL, Moss DM, Nayak V, Srinivasamoorthy G, Sheth M, Arrowood MJ. 2015. Draft genome sequences from *Cyclospora cayetanensis* oocysts purified from a human stool sample. *Genome Announc* 3(6):e01324-15. doi:10.1128/genomeA.01324-15.

**Copyright** © 2015 Qvarnstrom et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Yvonne Qvarnstrom, [bvp2@cdc.gov](mailto:bvp2@cdc.gov).

*Cyclospora cayetanensis* is a coccidian parasite that causes cyclosporiasis. Humans become infected by ingesting food or water contaminated with mature (sporulated) oocysts. The most common symptom of infection is watery diarrhea (1). Cyclosporiasis is not thought to be endemic in the United States: the two established risk factors for U.S. cases are international travel to regions where cyclosporiasis is endemic and the consumption of contaminated fresh produce imported from such regions (1, 2). U.S. foodborne outbreaks have been detected almost every year since the mid-1990s (<http://www.cdc.gov/parasites/cyclosporiasis/outbreaks/foodborneoutbreaks.html> [1]). However, outbreak investigations and studies of the epidemiology of cyclosporiasis have been hampered by multiple factors, including the lack of laboratory methods for strain subtyping and for sensitive detection of oocysts in food and environmental samples. One approach for developing improved laboratory methods is to use genomics to identify potential genetic markers for parasite detection and subtyping. Since very limited genetic information is available, the objective of this study was to obtain whole-genome sequences of *C. cayetanensis*.

No methods are available for propagating this parasite in the laboratory (3). A human stool sample collected during an outbreak investigation in 2001 was selected for this study because it had a large number of oocysts ( $>10^7$ ). The stool sample had been stored in 2.5% potassium dichromate at 4°C since the time of collection and was identified as positive for *Cyclospora* sp. by UV fluorescence microscopy (4). The oocysts were purified from stool components using discontinuous sucrose and cesium chloride gradients, as described elsewhere (5). The highly autofluorescent oocysts were further purified by flow cytometry sorting (FACSAria III; BD Biosciences) and then by treatment with 50% household bleach for 10 min on ice. Genomic DNA was released from oocysts by 15 cycles of freeze-thaw incubations and purified using the DNeasy blood & tissue kit (Qiagen). Four aliquots of genomic DNA were sheared to average sizes of 453, 584, 862, and 1,054 bp, respectively, in an M220 ultrasonicator (Covaris). About 10 ng of each sheared DNA sample was used for library construction

for Illumina sequencing using Ovation Ultralow library systems V2 (NuGEN). The barcoded libraries were subjected to sequencing using Illumina MiSeq reagent kits v2 (500 cycles) and v3 (600 cycles). A total of 147,760,794 reads were trimmed and assembled *de novo* using CLC Genomics Workbench 7.5.1 (CLC bio). The three best assemblies (the ones with the longest contig, the largest  $N_{50}$  value, and the lowest number of contigs, respectively) were merged using GAM-NGS (6) to enhance assembly contiguity and accuracy. The resulting assembly yielded 865 contigs with a total length of 44,563,857 bases, a G+C content of 51.9%, an  $N_{50}$  value of 187,023 bases, and a longest contig of 1,465,153 bases. The assembly included a full-length mitochondrial genome of 6,273 bases (contig 438) and two contigs comprising the apicoplast genome of 24,124 (contig 312) and 5,557 bases (contig 451), respectively.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [LIGJ00000000](https://www.ncbi.nlm.nih.gov/nuclink/LIGJ00000000). The version described in this paper is the first version, [LIGJ01000000](https://www.ncbi.nlm.nih.gov/nuclink/LIGJ01000000).

## ACKNOWLEDGMENTS

Financial support was provided by the CDC's Advanced Molecular Detection and Response to Infectious Disease Outbreaks Initiative. Fernanda Nascimento was supported by a fellowship (236608/2013-4) from the Brazilian National Council of Technological and Scientific Development (CNPq).

Alexandre J. da Silva, currently at CFSAN/FDA, contributed to the study design and the original proposal funded to support *C. cayetanensis* genome sequencing efforts. Andrew Huang, Lisle Garcia-Toledo, and others at the CDC Enteric Diseases Laboratory Branch provided technical support. Ryan Weil, Bin Hu, and Eldin Talundzic assisted with bioinformatic analysis.

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

## REFERENCES

- Herwaldt BL. 2000. *Cyclospora cayetanensis*: a review, focusing on the outbreaks of cyclosporiasis in the 1990s. *Clin Infect Dis* 31:1040–1057. [http://dx.doi.org/10.1086/314051](https://doi.org/10.1086/314051).

2. Hall RL, Jones JL, Herwaldt BL. 2011. Surveillance for laboratory-confirmed sporadic cases of cyclosporiasis—United States, 1997–2008. *MMWR Surveill Summ* 60:1–11.
3. Eberhard ML, Ortega YR, Hanes DE, Nace EK, Do RQ, Robl MG, Won KY, Gavidia C, Sass NL, Mansfield K, Gozalo A, Griffiths J, Gilman R, Sterling CR, Arrowood MJ. 2000. Attempts to establish experimental *Cyclospora cayetanensis* infection in laboratory animals. *J Parasitol* 86:577–582. [http://dx.doi.org/10.1645/0022-3395\(2000\)086\[0577:ATEECC\]2.0.CO;2](http://dx.doi.org/10.1645/0022-3395(2000)086[0577:ATEECC]2.0.CO;2).
4. Berlin OGW, Peter JB, Gagne C, Contreas CN, Ash LR. 1998. Autofluorescence and the detection of *Cyclospora* oocysts. *Emerg Infect Dis* 4:127–128. [http://wwwnc.cdc.gov/eid/article/4/1/98-0121\\_article](http://wwwnc.cdc.gov/eid/article/4/1/98-0121_article).
5. Arrowood MJ, Donaldson K. 1996. Improved purification methods for calf-derived *Cryptosporidium parvum* oocysts using discontinuous sucrose and cesium-chloride gradients. *J Eukaryot Microbiol* 43:89S. <http://dx.doi.org/10.1111/j.1550-7408.1996.tb05015.x>.
6. Vicedomini R, Vezzi F, Scalabrin S, Arvestad L, Policriti A. 2013. GAM-NGS: genomic assemblies merger for next generation sequencing. *BMC Bioinformatics* 14(Suppl 7):S6. <http://dx.doi.org/10.1186/1471-2105-14-S7-S6>.