Journal of Nematology 49(2):127–128. 2017.

The Draft Genome of *Globodera ellingtonae*

Wendy S. Phillips,[1] Dana K. Howe,[2] Amanda M. V. Brown,[2] Sebastian Eves-van den Akker,[3] Levi Dettwyler,[2] Amy B. Peetz,[1] Dee R. Denver,[2] and Inga A. Zasada[1]

Abstract: Globodera ellingtonae is a newly described potato cyst nematode (PCN) found in Idaho, Oregon, and Argentina. Here, we present a genome assembly for *G. ellingtonae*, a relative of the quarantine nematodes *G. pallida* and *G. rostochiensis*, produced using data from Illumina and Pacific Biosciences DNA sequencing technologies.
Key words: genome, *Globodera ellingtonae*, potato cyst nematode.

PCN are a global problem estimated to cause losses of 9% of total potato production worldwide (Turner and Rowe, 2006). The recent genome projects for *Globodera pallida* and *G. rostochiensis* contribute to the understanding and control of these economically important plant-parasitic nematodes (Cotton et al., 2014; Eves-van den Akker et al., 2016). A recently described PCN, *Globodera ellingtonae* (Handoo et al., 2012), is a close relative of *G. rostochiensis* and *G. pallida*. Here, we report on the genome sequencing and assembly of an Oregon strain of *G. ellingtonae*, likely the result of a substantial population bottleneck, providing a new avenue for understanding the evolution and biology of PCN.

We assembled the genome of *G. ellingtonae* using HiSeq and MiSeq reads and the de novo assembler Allpaths-LG (Gnerre et al., 2011), followed by gap filling with PBJelly using PacBio reads (English et al., 2012). Sequencing libraries were constructed from *G. ellingtonae* DNA extracted using a Qiagen DNeasy Blood & Tissue Kit (Valencia, CA) from second-stage juveniles hatched from several hundred cysts ($n = 200$) collected at Powell Butte, Oregon. MiSeq sequencing was performed for 301 × 2 cycles (paired-end); HiSeq sequencing was performed for 100 × 2 cycles (paired-end) for two mate pair libraries (~3 and ~8 kb inserts); and PacBio sequencing was performed on eight SMRTcells with 3 to 20 kb libraries. The MiSeq and PacBio DNA sequencing used the same original DNA extraction; the HiSeq run used a separate extraction performed later with ~500 cysts. Quality filtering and trimming of the Illumina reads were performed using the program Trimmomatic (Bolger et al., 2014); PacBio reads were corrected with MiSeq reads using the program WGS module pacBioToCA (Koren et al., 2012). To fulfill the Allpaths-LG assembly program requirement of short insert paired-end reads, single unpaired MiSeq reads were split into two overlapping 165 bp fastq sequences to mimic paired-end reads from a library with ~270 bp inserts. Percent reads used in the assembly were 69.9%, 50.6%, and 36.3% for the MiSeq, 3 kb HiSeq, and 8 kb HiSeq libraries, respectively, with final sequence coverages of 67.6, 50.5, and 27.0 for each. We screened for contamination using a blastn search of the GenBank microbe "Representative genomes" database, resulting in the removal of eight scaffolds, all <2 kb.

The final assembly contained 2,248 scaffolds, a number smaller than those resulting for the published assemblies of *G. rostochiensis* (4,377 scaffolds) and *G. pallida* (6,873 scaffolds). The final *G. ellingtonae* assembly totaled 119,060,168 bp, similar to the *G. rostochiensis* and *G. pallida* assemblies (95.9 and 124.6 Mb, respectively). The longest *G. ellingtonae* scaffold was 2.8 Mb (compared to 0.7 Mb for *G. rostochiensis* and 0.6 Mb for *G. pallida*), and 15 scaffolds were >1 Mbp. The final *G. ellingtonae* assembly had a scaffold N50 of 360 kb (89 kb for *G. rostochiensis* and 122 kb for *G. pallida*), contained 11.7% gaps (5% for *G. rostochiensis* and 17% for *G. pallida*), and had a GC content of 37% (38% for *G. rostochiensis* and 37% for *G. pallida*).

To assess completeness of the *G. ellingtonae* assembly, the program CEGMA (Parra et al., 2007) was used to search for 248 core eukaryotic genes (CEG). At least partial transcripts were detected for 239 (96%) of the CEGs, with complete transcripts for 229 (92%). These values were similar to those for the *G. rostochiensis* genome (96% partial and 94% complete), and higher than those for *G. pallida* (81% partial and 74% complete). We further compared the *G. ellingtonae* genome to that of *G. rostochiensis* by using the 14,309 predicted proteins from the latter species as queries in blat searches against the former species' genome. Using an identity cutoff of 60%, we found hits for 14,104 (98.6%) of the proteins; 13,946 (97.4%) hits were found using a more conservative cutoff of 80%. The *G. ellingtonae* genome provides a valuable resource for study of the evolution in the PCN lineage.

GenBank accession numbers: The raw DNA sequence data and genome assembly were deposited at GenBank under BioSample no. SAMN04393202.

Received for publication February 1, 2017.
[1]USDA-ARS Horticultural Crops Research Laboratory, 3420 NW Orchard Avenue, Corvallis, OR 97330.
[2]Department of Integrative Biology, Oregon State University, 3029 Cordley Hall, OR 97331.
[3]Division of Plant Sciences, College of Life Sciences, University of Dundee, Dundee, DD1 5EH, UK.
We thank Mark Dasenko for running the Illumina libraries and constructing the mate-pair libraries and Dominik Laetsch for advice on analyses. This research was partially funded by the Northwest Potato Research Consortium, USDA-ARS, and USDA-APHIS.
E-mail: denvedee@oregonstate.edu.
This paper was edited by Erik Ragsdale.

Literature Cited

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., Tsai, I. J., Beasley, H., Blok, V., Cock, P. J., Eves-van den Akker, S., Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J. E., Zarowiecki, M., Berriman, M., Jones, J. T., and Urwin, P. E. 2014. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. Genome Biology 15:R43.

English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D. M., Reid, J. G., Worley, K. C., and Gibbs, R. A. 2012. Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One. 2012:7.

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M. O., Boyes, I., Marvin, J. M., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., and Rott, M. 2016. The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. Genome Biology 17:124.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., Berlin, A. M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E. S., and Jaffe, D. B. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences 108:1513–1518.

Handoo, Z. A., Carta, L. K., Skantar, A. M., and Chitwood, D. J. 2012. Description of *Globodera ellingtonae* n. sp (Nematoda: Heteroderidae) from Oregon. Journal of Nematology 44:40–57.

Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., and Adam, M. P. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. Nature Biotechnology 30:693–700.

Parra, G., Bradnam, K., and Korf, I. 2007. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genornes. Bioinformatics 23:1061–1067.

Turner, S. J., and Rowe, J. A. 2006. Cyst nematodes. Pp. 90–122 *in* R. N. Perry and M. Moens, eds. Plant nematology. Wallingford, Oxfordshire: CAB International.