Draft Transcriptome of *Globodera ellingtonae*

Wendy S. Phillips,[1] Sebastian Eves-van den Akker,[2] and Inga A. Zasada[1]

Abstract: Globodera ellingtonae is a newly described cyst nematode found in Idaho, Oregon, and Argentina. Here we present the first transcriptome assembly of *G. ellingtonae*, providing a valuable resource for comparing the evolution of expressed genes between potato cyst nematode species.

Key words: Globodera ellingtonae, potato cyst nematode, transcriptome.

In 2012, *Globodera ellingtonae*, a potato cyst nematode (PCN), was described from populations collected in Idaho and Oregon (Handoo et al., 2012). Given that assembled transcriptomes are available for *G. rostochiensis* and *G. pallida* (Cotton et al., 2014; Eves-van den Akker et al., 2016), the addition of a transcriptome for *G. ellingtonae* will provide a unique perspective on the evolution of expressed genes within this genus.

To assemble a reference transcriptome for *G. ellingtonae*, RNA was extracted from sedentary second-, third-, and fourth-stage juveniles, eggs, preinfective juveniles, and adult males and females using the Quick-RNA MicroPrep kit (Zymo Research, Irvine, CA). A library was created by pooling ~100 ng RNA from each life stage, using WaferGen PrepX PolyA and PrepX RNA-seq kits (WaferGen Biosystems, Fremont, CA), and run on an Illumina MiSeq to generate 29,876,195 75 bp, paired-end reads. After quality trimming to a minimum Q score of 30, the paired reads were assembled using default options in Trinity v2.1.1 (Grabherr et al., 2011). Isoforms with abundances <5% of the parent gene were removed, yielding an initial assembly with 30,541 transcripts totaling 20,622,916 bases. Using a stringent approach to eliminate potential contamination, transcripts were removed for which top blastn (-nt, $<10^{-4}$) hits were to nonmetazoans, or that had no blastx ($<10^{-5}$) hits to peptides of *G. rotochiensis* or a custom peptide database from 42 nematode genomes (approximately 3,700 transcripts were removed). The final assembly totaled 18,980,927 bp, had an N50 of 1,028 bp and a GC content of 49.35%, and consisted of 17,355 gene clusters containing 22,977 genes for a total, including isoforms, of 26,318 assembled transcripts.

To assess completeness of the transcriptome, the program CEGMA was used to search for 248 core eukaryotic genes (CEGs). At least partial transcripts were detected for 203 (82%) of the CEGs, with complete transcripts for 179 (72%) (Parra et al., 2007). Using default parameter e-value cut off of 10, an HMMsearch (v3.1b1) of the transcriptome against 16,300 protein domains (Pfam v29.0 database; Finn et al., 2014) detected 8,764 pfam domains in 16,761 transcripts (http://hmmer.org). The majority of *G. rostochiensis* predicted proteins (86% of the 14,309) had a hit in a tblastn search ($<10^{-5}$) to *G. ellingtonae* transcripts. Blastx ($<10^{-4}$) searches against the distantly related *Caenorhabditis elegans* proteome resulted in fewer hits: 15,258 (58%) *G. ellingtonae* transcripts hit *C. elegans* proteins. The *G. ellingtonae* transcriptome provides a valuable resource for study of the evolutionary dynamics of expressed genes in the PCN lineage.

Nucleotide sequence accession numbers: Raw sequence reads are available under the Short Read Archive (SRA) accession no. SRR3162514. The transcriptome was deposited in the GenBank Transcriptome Shotgun Assembly (TSA) Sequence Database under accession no. GEZD00000000.

Literature Cited

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. A., 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.

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., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P. J., Phillips, W., Henrissat, B., Urwin, P. E., Blaxter, M., and Jones, J. T. 2016. The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. Genome Biology 17:124.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. 2014. Pfam: the protein families database. Nucleic Acids Research 42:D222–D230.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644–652.

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.

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