exeley

Article | DOI: 10.21307/jofnem-2019-027

JOURNAL OF NEMATOLOGY

e2019-27 | Vol. 51

The draft genome of *Ditylenchus dipsaci*

Benjamin Mimee[1,*], Etienne Lord[1], Pierre-Yves Véronneau[1], Rick Masonbrink[2], Qing Yu[3] and Sebastian Eves-van den Akker[4]

[1]Agriculture and Agri-Food Canada, St-Jean-sur-Richelieu Research and Development Centre, St-Jean-sur-Richelieu, J3B 3E6 QC, Canada.

[2]Iowa State University, Ames, IA.

[3]Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, K1A 0C6 Ottawa, ON, Canada.

[4]Department of Plant Sciences, University of Cambridge, Cambridge.

*E-mail: benjamin.mimee@canada.ca

This paper was edited by Dee Denver.

Received for publication November 27, 2018.

Abstract

Ditylenchus dipsaci is a devastating pest to many crops worldwide. We present the first genome sequence for this species, produced with PacBio sequencing and assembled with CANU.

Key words

Canu, Genome assembly, PacBio, Stem and bulb nematode.

The stem and bulb nematode, *Ditylenchus dipsaci*, is a plant-parasitic nematode affecting over 500 plant species with major destruction occurring in onion, garlic, beet, strawberry, broad bean, potato, alfalfa, and oat (Sturhan and Brzeski, 1991). It is a quarantine pest in Europe (EPPO, 2004) and well known in garlic production where it can cause losses of up to 90% (Abawi and Moktan, 2010). The species exhibits a variety of host preferences for which biological groups were proposed (Sturhan and Brzeski, 1991). In this study, we report the first genome assembly of *D. dipsaci*, a resource that will improve our understanding of this important parasite.

A population of *D. dipsaci* (E-105) was obtained from the *Canadian National Collection of Insects, Arachnids, and Nematodes* and originally isolated from garlic in Northern Ontario (Qiao et al., 2013). The nematodes were multiplied *in vitro* on pea sprout inoculated with only 15 mixed-stages individuals, as described in Poirier et al. (2019), to reduce heterozygosity. DNA was extracted from ~7,500 mixed-stage *D. dipsaci,* using the Qiagen DNeasy Blood & Tissue Kit (Valencia, CA). *MUGQIC* (Montréal, Canada) performed the library preparation (sheared large insert) and sequencing on a PacBio RSII instrument using 12 SMRT cells yielding 941,946 long reads at a mean size of 10,207 bp for an estimated coverage of 42×. Two assemblies were generated using CANU v1.7 (Koren et al., 2017), one with default parameters and one optimized for high heterogeneity (corMhapSensitivity=high, corMinCoverage=0, and correctedErrorRate=0.105). These pre-assemblies were merged using Quickmerge (Chakraborty et al., 2016) and decontaminated using Blobtools v1.0 (Laetsch and Blaxter, 2017). Before scaffolding, the assembly contained 3,559 contigs with a N50 of 152 kb. Contigs were then scaffolded using MeDuSa v1.6 (Bosi et al., 2015) with the genomes of *Caenorhabditis elegans* (GCA_000002985.3; The C. elegans Sequencing Consortium, 1998), *D. destructor* (GCA_001579705.1; Zheng et al., 2016), *Globodera pallida* (GCA_000724045.1; Cotton et al., 2014), *G. rostochiensis* (GCA_900079975.1; Eves-van den Akker et al., 2016), and *Meloidogyne incognita* (GCA_900182535.1; Abad et al., 2008). Scaffolding was further improved using Rascaf (Song et al., 2016) and 13,555,476 PE RNA-seq reads (Phred score > 20), obtained from mixed-stage *D. dipsaci* (Illumina HiSeq 2×100 bp, TruSeq Library Prep Kit).

Then, Purge Haplotigs (Roach et al., 2018) resolved highly polymorphic regions and reduced artifactual duplication.

The *D. dipsaci* genome assembly is comprised of 1,394 scaffolds, a largest scaffold of ~3.6 Mb, 22 scaffolds >1 Mb, and a N50 of 287 kb. The assembly is 227,234,012 bp (227.2 Mb), double the size of the *D. destructor* assembly (111.1 Mb). We used flow cytometry on *D. dipsaci* nuclei to investigate this discrepancy, revealing a haploid genome size of 160 to 170 Mb. Thus, obtaining 227 Mb with heterozygous material is not surprising. The GC content of the *D. dipsaci* assembly was 37.5% which is comparable to *D. destructor* (36.8%), *C. elegans* (35.4%), *G. rostochiensis* (38.1%), and *G. pallida* (36.7%). A completeness assessment with BUSCO v 3.0.2 (Simão et al., 2015) using the Nematoda_odb9 database revealed 537 complete sequences (506 single-copy, 31 duplicated), 133 fragmented, and 312 missing on a total of 982 BUSCO genes searched in *D. dipsaci.* The number of complete genes was higher than *G. pallida* (431) but lower than *D. destructor* (749). Gene prediction was performed with Braker 2.1.0 (Hoff et al., 2015) on a genome masked with Repeatmodeler/Repeatmasker 4.0.7 (Chen, 2004), leaving simple repeats unmasked. Transcript information was relayed to Braker via HISAT2 2.1.0 alignments of *D. dipsaci* RNA-seq and exonerate 2.4.0 aligned *Ditylenchus destructor* proteins to produce a gene annotation of 26,428 putative genes. On average, these genes contained seven exons (max = 114) with a mean length of 120 bp (from 2 to 9,194 bp) and six introns (max = 113) with a mean length of 515 bp (from 36 to 27,260 bp).

GenBank accession numbers: The raw sequences and genome assembly files are available under the NCBI BioProject PRJNA498219.

Acknowledgments

The authors thank Sandra Poirier and Nathalie Dauphinais for technical assistance with nematode rearing and DNA extraction. This work was funded by Agriculture and Agri-Food Canada.

References

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Ségurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., G Maule, A., Quesneville, H., Rosso, M.-N., Schiex, T., Smant, G., Weissenbach, J. and Wincker, P. 2008. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nature Biotechnology 26:909–915.

Abawi, G. S. and Moktan, K. 2010. Bloat nematode problem on garlic: symptoms, distribution, and management guidelines. Department of Plant Pathology and Plant-microbe Biology, Geneva: Cornell University.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M. F., Lió, P., Crescenzi, P., Fani, R. and Fondi, M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–2451.

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. and Emerson, J. J. 2016. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Research 44:1–12.

Chen, N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics 5:4–10.

Cotton, J. A., et al. 2014. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. Genome Biology 15:1–17.

EPPO. 2004. Data Sheets on Quarantine Pests: *Ditylenchus dipsaci*. E. a. M. P. P. Organization and I. S. Compendium, 90/399003.

Eves-van den Akker, S., et al. 2016. The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. Genome Biology 17:1–23.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. 2015. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 27:722–736.

Laetsch, D. R. and Blaxter, M. L. 2017. BlobTools: Interrogation of genome assemblies. F1000Research, 6:1287.

Poirier, S., Dauphinais, N., Van der Heyden, H., Véronneau, P. Y., Bélair, G., Gravel, V. and Mimee, B. 2019. Host range and genetic characterization of *Ditylenchus dipsaci* populations from eastern Canada. Plant Disease, 103:456–460.

Qiao, Y., Zaidi, M., Badiss, A., Hughes, B., Celetti, M. and Yu, Q. 2013. Intra-racial genetic variation of *Ditylenchus dipsaci* isolated from garlic in Ontario as revealed by random amplified polymorphic DNA analysis. Canadian Journal of Plant Pathology 35:346–353.

Roach, M. J., Schmidt, S. A. and Borneman, A. R. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19:460.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Song, L., Shankar, D. S. and Florea, L. 2016. Rascaf: improving genome assembly with RNA sequencing data. The Plant Genome 9:1–12.

Sturhan, D. and Brzeski, M. W. 1991. Stem and bulb nematodes, *Ditylenchus* spp. in Manual of agricultural nematology by Marcel Dekker. New York, NY: W.R. Nickle, pp. 423–464.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012–2018.

Zheng, J., Peng, D., Chen, L., Liu, H., Chen, F., Xu, M., Ju, S., Ruan, L. and Sun, M. 2016. The *Ditylenchus destructor* genome provides new insights into the evolution of plant parasitic nematodes. Proceedings of the Royal Society B 283:20160942.