JOURNAL OF NEMATOLOGY

Research Paper | DOI: 10.2478/jofnem-2023-0031

e2023-1 | Vol. 55

Nematode Genome Announcement: A Draft Genome of Seed Gall Nematode, *Anguina tritici*

Ashish Kumar Singh[1,2], Antara Das[3], Ila Joshi[3], Manish Kumar[1], Deshika Kohli[3], Pankaj[1], Kishor Gaikwad[3], Pradeep Kumar Jain[3] and Anil Sirohi [1,*]

[1]Division of Nematology, ICAR-Indian Agricultural Research Institute, Pusa Campus, New Delhi-110012, India.

[2]Division of Crop Protection, ICAR-Vivekananda Parvatiya Krishi Anusandhan Sansthan, Almora, Uttarakhand 263601, India.

[3]ICAR-National Institute for Plant Biotechnology, Pusa Campus, New Delhi, 110012, India.

*E-mail: anilsirohi@yahoo.com

This paper was edited by Erik C Andersen.

Received for publication December 8, 2022.

Abstract

Anguina tritici is the first plant-parasitic nematode described in literature, dating back to the year 1743. It is responsible for causing earcockle (seed gall) and tundu diseases in wheat and rye. Notably, this nematode has been observed to survive in an anhydrobiotic state for up to 32 years within wheat seed galls. These exceptional characteristics have inspired the sequencing of the *A. tritici* genome. In this study, we present the initial draft genome of *A. tritici*, obtained using the Illumina MiSeq platform with coverage of 60-fold. The genome is estimated to have a size of 164 Mb and comprises 39,965 protein-coding genes, exhibiting a GC content of 39.1%. The availability of this genome data will serve as a foundation for future functional biological investigations, particularly for genes whose functions remain unknown to this day.

Keywords

Anguina tritici, earcockle, tundu disease, genome sequence, Illumina MiSeq

At least four distinct major lineages of plant parasites have emerged within the species-rich and trophically diverse phylum Nematoda (Quist et al., 2015). *A. tritici*, commonly known as the seed gall nematode, is a plant parasitic nematode (PPN) that infects the seeds of wheat and barley, leading to the development of the earcockle disease. This nematode also acts as a vector for the bacteria *Clavibacter tritici*, causing tundu disease in their association (Swarup and Gupta, 1971). According to the phylogenetic analysis conducted by van Megen et al. (2009), *Anguina tritici* is classified within the Anguinidae family, along with its closely related species *Ditylenchus dipsaci* and *D. destructor*. In the past decade, the genomes of several parasitic nematodes have been sequenced, enabling comparative genomic analyses that shed light on the gain or loss of novel genes among nematode species (Wasmuth et al., 2008). As more genome sequences become available, researchers will be able to identify species-specific peculiarities in biological characteristics.

The biology of *A. tritici* differs from other PPNs, which usually feed on or in the roots of the crop plants but do not infest the seeds. *A. tritici* is an obligate parasite specifically targeting the seeds of its host plant wheat and barley, unlike the rice white tip nematode, *Aphelenchoides besseyi*, which also survives in an anhydrobiotic state but in dry tissues, under hulls of rice grains. *A. tritici* exhibits the remarkable ability to survive for several decades in an anhydrobiotic state without losing its parasitic capability (Limber, 1973) but *A. besseyi* is reported to survive for a maximum of three years without losing its parasitic ability (Cralley, 1949; Yoshi and Yamamoto, 1950) Among plant parasitic nematodes, *A. tritici* exhibits notable variability in certain biological traits, distinguishing it from the other PPN species and making it an attractive model system for investigating fundamental questions related to survival, ageing, seed parasitism, bacterial associations, and more. Currently, limited genomic information is available regarding plant-parasitic nematodes that specifically target crop seed.

Open Access. Published by Sciendo. BY-NC-ND

© 2023 Singh et al. This is an Open Access article licensed under the Creative Commons CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/

Recently, the genome of *Aphelenchoides besseyi*, a nematode that feeds on the growing tips of rice seedlings, causing the white-tip disease, has been published (Lai et al., 2023; Ji et al., 2023). However, unlike *A. tritici*, it does not enter the grain but remains localized in the husk part of the paddy seed. The genome sequences of two closely related species of *Anguina tritici*, namely *Ditylenchus dipsaci* (Mimee et al., 2019) and *D. destructor* (Zheng et al., 2016) are also available. These genome sequences provide valuable resources for comparative analysis and serve as important references for understanding the genomic features of *A. tritici* in a broader context. The objective of this study was to present the draft genome of the seed gall nematode, *A. tritici*. To propagate the *A. tritici*, cockled seeds were crushed and placed in a modified Baermann funnel for the extraction of second-stage juveniles (J2s). The extracted J2s were then inoculated onto greenhouse-grown wheat cultivar 'PBW 343' for multiple generations to generate an inbred population of the nematode. Before DNA extraction, J2s were surface sterilized using M9 buffer and nematode sterilization buffer to remove any contaminants (Joshi et al., 2019). High-quality genomic DNA was extracted using the modified fast CTAB method (Thomas et al., 1997). The isolated high-quality DNA was utilized for the preparation of a genomic DNA sequencing library using the TruSeq DNA sample preparation kit from Illumina technologies, USA by following the manufacturer's protocol. The quality of DNA libraries was assessed using the Bioanalyzer 2100 from Agilent technologies, USA. Paired-end sequencing of the genomic library was performed on the Illumina MiSeq (2X300) platform. The genome size was estimated using the K-mer value of the Abyss Version 1.3.7 (Simpson et al., 2009). Paired-end files were merged using PandaSeq 2.5 (Masella et al., 2012), and the reads were subjected to quality checking using FastQC software v3.0 (Andrew, 2012). High-quality reads with a Phred score of ≥20, after trimming adapters using trimmomatic version 0.39 (Bolger et al., 2014), were retained. *De novo* assembly was carried out using Abyss Version 1.3.7 (Simpson et al., 2009) at different K-mers (45, 50, 55, and 61). The quality assessment of the assembled genome was performed using BUSCO version 3.0.2 (Robert et al., 2017). Protein-coding genes were predicted using the *ab initio* approach of AUGUSTUS version 3.0.3 (Stanke et al., 2004) with default parameters. Predicted genes with a length of less than 400 bp were excluded, and the remaining genes were annotated against the NCBI redundant nucleotide database using RAPsearch2 and an in-house perl script (Yuzhen et al., 2011).

The Illumina MiSeq sequencing strategy generated a total of 11,065,381,294 bases, with a read count of 40,210,848. The GC percentage, Q20 percentage, and Q30 percentage were calculated as 39.1%, 92.73%, and 84.08%, respectively. By merging the paired-end reads using the Pandaseq program, a total of 7,938,027,993 bases and 18,913,374 read counts were obtained, with a GC percentage of 39.1%. The draft genome assembly of *A. tritici* resulted in a size of up to 164 Mb with a GC content of 39.1%. The genome sequence data has been submitted to GenBank with the following identifiers: BioProject ID: PRJNA802691, BioSample: SAMN25565353, and accession ID: JAKMXJ000000000. The genome achieved a total coverage level of 60-fold. Quantitative assessment of assembly completeness was performed using BUSCO, revealing C: 36.6% [S: 35.5%, D: 1.1%], F: 14.8%, M: 48.6%, and n: 982 (Table 1). A total of 44,954 protein coding genes were predicted using the *ab initio* Augustus version 3.0.3. After filtering genes with a length below 400 base pairs, a total of 39,965 predicted genes were identified, resulting in an overall gene density of 26.5784. The average gene length was 1,347 bp, while the average exon and intron lengths were 168 bp and 280 bp, respectively. Of the 41,746 predicted transcripts, 62% were successfully annotated with functional annotations against the NCBI database using RAP search and an in-house perl script. However, 14,856 (37.18%) genes remained unannotated in the NCBI database. In comparison with closely related species, the genome size of *Anguina tritici* (164 Mb) falls between that of *Ditylenchus dipsaci* (227.2 Mb) (Mimee et al., 2019) and *Ditylenchus destructor* (112 Mb) (Zheng et al., 2016) indicating that *Anguina tritici* has a moderately sized genome in comparison to the other two species. The GC content, shows slight variations among the three species. *Anguina tritici* has the highest GC content, indicating a relatively higher abundance of G and C nucleotides in its genome compared to the *Ditylenchus dipsaci* (37.5) and *Ditylenchus destructor* (36.6). *Anguina tritici* has the highest number of coding genes (39,965) among the three species, indicating a more complex and diverse genetic repertoire. *Ditylenchus dipsaci* has an intermediate number of genes (26,428), while *Ditylenchus destructor* has the fewest genes (13,938). These variations may reflect evolutionary adaptations and functional differences between these nematode species.

Broadly, the genome size of *A. tritici* was determined to be 164 Mb, which is larger than most of the published genomes of PPNs, except

Table 1. Characteristics of *A. tritici* genome.

Number of contigs	843604
Assembled genome size	164398910 (164 Mb)
N50	3232
Number of gene models	39965
Median intron, bp	65
G+C %	39.1%
Gene density	26.5784
Average gene size (bp)	1347
Average CDS length (bp)	168.88
Average number of exon/genes	3.18
Average intron length (bp)	280
BUSCO	C: 36.6% [S: 35.5%, D: 1.1%], F: 14.8%, M: 48.6%, n: 982

for *Meloidogyne arenaria*, *Rotylenchulus reniformis*, and which are 284 Mb and 380 Mb, respectively (Nyaku et al., 2014; Sato et al., 2018). The GC content was measured to be 39.1%, which is higher than most PPNs, except for *Bursaphelenchus xylophilus* (Kikuchi et al., 2011). From the *A. tritici* genome, a total of 39,965 protein-coding genes were predicted. This number is higher than most of the PPN genomes sequenced, except for *Meloidogyne javanica*, *Meloidogyne floridensis*, and *Meloidogyne arenaria* (Blanc-Mathieu et al., 2017). On an average, three exons per gene were observed, compared to two exons per gene in *Pratylenchus coffeae* (Burke et al., 2015) and four in most PPNs. The *A. tritici* genome exhibits characteristics of a compact parasitic genome, similar to other nematodes, with a low number of small introns and a high number of repeats. Out of the 41,746 predicted transcripts, 62% were successfully annotated with functional annotations. These unannotated genes may possess unique functions in the *A. tritici* genome since they do not show significant matches in the NCBI database. Their validation could shed light on their species-specific roles. The *A. tritici* genome will be valuable in unravelling the mechanisms involved in anhydrobiosis and the genes responsible for parasitizing the seed parts of plants, an area of nematology that has been less explored. Integrating RNAseq data in the annotation process could enhance gene prediction and annotation accuracy, thereby improving the overall quality and completeness of the final genome assembly. Future studies should consider incorporating RNAseq data to further enhance the understanding of the *A. tritici* genome.

Acknowledgments

We acknowledge funding, facilities, and support provided by ICAR-IARI and ICAR-NIPB, New Delhi to carry out the study.

Literature Cited

Swarup, G., and Gupta P. 1971. On the ear-cockle and tundu diseases of wheat. II. Studies on *Anguina tritici* (Steinbuch, 1799) Filipjev, 1936 and *Corynebacterium tritici* (Hutch.) Burkh. Indian Phytopathology 24:359–365.

Wasmuth, J., Schmid, R., Hedley, A., and Blaxter, M. 2008. On the extent and origins of genic novelty in the phylum Nematoda. PLOS Neglected Tropical Diseases 2(7):e258.

Limber, D.P. (1973). Notes on the longevity of Anguina tritici (Steinbuch, 1799) Filipjev, 1936, and its ability to invade wheat seedlings after thirty-two years of dormancy. Proceedings of the Helminthological Society, Volume 40, Number 2, July 1973.

Joshi, I., Kumar, A., Singh, A. K., Kohli, D., Raman, K. V., Sirohi, A., Chaudhury, A., and Jain, P. K. 2019. Development of nematode resistance in Arabidopsis by HD-RNAi-mediated silencing of the effector gene *Mi-msp2*. Scientific Reports 9:17404. https://doi.org/10.1038/s41598-019-53485-8.

Thomas, J. C., Khoury, R., Neeley, C. K., Akroush, A. M., and Davies, E. C. 1997. A fast CTAB method of human DNA isolation for polymerase chain reaction applications. Biochemical Education 25(4):233–235.

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., and Neufeld, J. D. 2012. PANDAseq: Paired-end Assembler for Illumina sequences. BMC Bioinformatics 13(1):31.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., and Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Research 19:1117–1123.

Robert, M. W., Mathieu S., Felipe, A., Simão, M., Manni, P. I., Guennadi, K., Evgenia, V. K., and Evgeny M. Z. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution 35(3):543–548.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Research 32:W309–W312.

Yuzhen, Y., Choi, J. H., and Haixu, T. 2011. RAPSearch: a fast protein similarity search tool for short reads. BMC Bioinformatics 12:159.

Sato, K., Kadota, Y., Gan, P., Bino, T., Uehara, T., Yamaguchi, K., and Shirasu, K. 2018. High-quality genome sequence of the root-knot nematode *Meloidogyne arenaria* genotype A2-O. Genome Announcements 6(26).

Nyaku, S. T., Sripathi, V. R., Kantety, R. V., Cseke, S. B., Buyyarapu, R., Mc Ewan, R., and Sharma, G. C. 2014. Characterization of the reniform nematode genome by shotgun sequencing. Genome 57(4): 209–221.

Kikuchi, T., Cotton, J.A., Dalzell, J.J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takanashi, T., Tsai, I.J., Assefa, S.A., Cock, P.J. and Otto, T.D. 2011. Genomic insight into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. PLoS Pathogen 7(9), p.e1002219.

Quist, C. W., Smant, G., and Helder, J. 2015. Evolution of plant parasitism in the phylum Nematoda. Annual Review of Phytopathology 53:289–310. doi: 10.1146/annurev-phyto-080614-120057. PMID: 26047569

Mimee, B., Lord, E., Véronneau, P. Y., Masonbrink, R., Yu, Q., and Akker, S. E. D. 2019. The draft genome of *Ditylenchus dipsaci*. Journal of Nematology 51:1–3. doi: 10.21307/jofnem-2019-027. PMID: 31132003; PMCID: PMC6929663

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.M., Da Rocha, M., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.F. and Kozlowski, D.K. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. PLoS Genetics 13(6):e1006777. https://doi.org/10.1371/journal.pgen.1006777

Burke, M., Scholl, E. H., Bird, D. M., Schaff, J. E., Colman, S. D., Crowell, R., Diener, S., Gordon, O., Graham, S., Wang, X., Windham, E., Wright, G. M., and Opperman, C. H. 2015. The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. Nematology 17(6):621-637. doi: https://doi.org/10.1163/15685411-00002901

Zheng, J., Peng, D., Chen, L., Liu, H., Chen, F., Xu, M., and Sun, M. 2016. The *Ditylenchus destructor* genome provides new insights into the evolution of plant parasitic nematodes. Proceedings of the Royal Society B: Biological Sciences 283(1835):20160942.

Lai, C.K., Lee, Y.C., Ke, H.M., Lu, M.R., Liu, W.A., Lee, H.H., Liu, Y.C., Yoshiga, T., Kikuchi, T., Chen, P.J. and Tsai, I.J. 2023. The *Aphelenchoides* genomes reveal substantial horizontal gene transfers in the last common ancestor of free-living and major plant parasitic nematodes. Molecular Ecology Resources 23(4):905–919.

Ji, H., Xie, J., Han, Z., Yang, F., Yu, W., Peng, Y., and Qing, X. 2023. Complete genome sequencing of nematode *Aphelenchoides besseyi*, an economically important pest causing rice white-tip disease. Phytopathology Research 5(1):1–13.

van Megen, H., van den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T., Holovachov, O., Bakker, J. and Helder, J. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology 11(6):927–950.

Cralley, E. M. 1949. White tip of rice. Phytopathology 39:5.

Yoshi, H., and Yamamoto, S. 1950. A rice nematode disease “Sencha Shingane Byo”. II. hibernation of *Aphelenchoides oryzae*. Journal of Faculty of Agriculture, Kyusha University 9:223–233.