The Genome of *Setaria digitata*: A Cattle Nematode Closely Related to Human Filarial Parasites

Kanchana S. Senanayake[1], Jonas Söderberg[2], Aleksei Põlajev[2], Maja Malmberg[2,3], Eric H. Karunanayake[1], Kamani H. Tennekoon[1], Sameera R. Samarakoon[1], Erik Bongcam-Rudloff[2,*], and Adnan Niazi [2]

[1]Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Sri Lanka

[2]SLU Global Bioinformatics Centre, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

[3]Section of Virology, Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden

*Corresponding author: E-mail: erik.bongcam@slu.se.

Accepted: January 21, 2020

Data deposition: This project has been deposited at European Nucleotide Archive (ENA) database under the accession number PRJEB13338.

Abstract

Here we present the draft genome sequence of *Setaria digitata*, a parasitic nematode affecting cattle. Due to its similarity to *Wuchereria bancrofti*, the parasitic nematode that causes lymphatic filariasis in humans, *S. digitata* has been used as a model organism at the genomic level to find drug targets which can be used for the development of novel drugs and/or vaccines for human filariasis. *Setaria digitata* causes cerebrospinal nematodiasis in goats, sheep, and horses posing a serious threat to livestock in developing countries. The genome sequence of *S. digitata* will assist in finding candidate genes to use as drug targets in both *S. digitata* and *W. bancrofti*. The assembled draft genome is ~90 Mb long and contains 8,974 genomic scaffolds with a G+C content of 31.73%.

Key words: *Setaria digitata*, cerebrospinal nematodiasis, setariosis, filariasis, helminth, genome, cattle parasite.

Introduction

Setaria digitata is a parasitic nematode found in the peritoneal cavity of cattle. Although nonpathogenic to their natural hosts, there have been instances where the larvae were observed in the cerebrospinal cavity (Tung et al. 2003), causing fatal paralysis. Transmission of infective larvae to aberrant hosts such as goats, sheep, cows and horses can cause a serious and often fatal neuropathological disorder commonly identified as cerebrospinal setariosis (Mohanty et al. 2000; Tung et al. 2003; Bazargani et al. 2008; Kaur et al. 2015; Shin et al. 2017). Due to its similarity to *Wuchereria bancrofti* with respect to morphology and histology (Decruse and Raj 1990) as well as antigenic properties (Dissanayake and Ismail 1980), *S. digitata* has been used as a model organism for the study of *W. bancrofti* in view of developing a vaccine against lymphatic filariasis (Madathiparambil et al. 2011; Perumal et al. 2016). A few genes from *S. digitata* have been characterized with the aim of identifying possible drug targets for *W. bancrofti* (Murugananthan et al. 2010; Rodrigo et al. 2013; Nagaratnam et al. 2014; Rodrigo et al. 2014). Although *W. bancrofti* is the major causative organism for lymphatic filariasis accounting for 90% of the cases (WHO 1992) and an estimated 120 million people in 72 countries were infected in 2010 (Dissanayake and Ismail 1980), very little is known about the molecular biology, biochemistry and immune mechanisms of this parasite.

The necessity of a model organism to study the biology of *W. bancrofti* has arisen because *W. bancrofti* grow in human lymph vessels and are therefore not easily isolated, and cannot be cultured in laboratory conditions. This is an impediment for testing new treatments for lymphatic filariasis in vitro (Murugananthan et al. 2010). Because the draft genome of the *W. bancrofti* is already publicly available (Small et al. 2019), the *S. digitata* genome will greatly facilitate comparative studies at the genetic level between the two nematodes and allow identification of genes that can be used as drug targets, which can then be tested in the laboratory. Furthermore, availability of the *S. digitata* genome will enable

drug development and vaccine production to eliminate or control *S. digitata* infections in abnormal hosts which cause serious economic loss in places where sheep and goat farming is a common livelihood (Tung et al. 2003; Nakano et al. 2007; Bazargani et al. 2008).

Materials and Methods

Collection of Adult *S. digitata* Worms

Adult *S. digitata* were collected from freshly slaughtered cattle at a nearby abattoir and immediately transported to the laboratory in sterile Hank's Buffered Salt Solution (Fisher Scientific, UK, Gibco Cat# 15420614). Adult worms were washed six times in sterile phosphate buffered saline (PBS) and used for genomic DNA extraction.

Extraction of Genomic DNA

Genomic DNA of adult *S. digitata* worms was extracted by phenol/chloroform extraction method and ethanol precipitated as described earlier (Nayak et al. 2012). Briefly, adult worms were lysed in 500 µl of lysis buffer (Tris–HCl 20 mM, ethylenediaminetetraacetic acid 50 mM, pH 8.0, SDS 0.5%, NaCl 100 mM, β mercaptoethanol 1%, v/v) containing 0.1 mg/ml proteinase K and incubated at 37 °C for 2 h. Then RNase (5 µg/ml) treatment was carried out for 2 h at 56 °C and DNA was isolated by phenol–chloroform–isoamyl alcohol extraction and ethanol precipitation. Total DNA amounts were quantified using the Qubit (Version 2.0) (ThermoFisherScientific, Waltham, MA) using the dsDNA assay. In parallel the integrity of the DNA was checked by gel electrophoresis on a 1.5% agarose gel for a qualitative assessment of DNA integrity.

Genome Sequencing

Sequencing libraries were constructed from the extracted DNA using the NexteraXT kit (Illumina, San Diego, CA) with small modifications; 10 cycles of amplification were used instead of 12, the samples were indexed as 10 different samples, libraries were normalized manually based on concentration measurements from Agilent High Sensitivity DNA Kit (2100 Bioanalyzer, Agilent Technologies, Palo Alto, CA). After library preparation the samples were pooled and sequencing was performed using the v3 600 cycles kit on the Illumina MiSeq platform (Illumina) at National Veterinary Institute in Uppsala.

Genome Assembly and Annotation

The raw sequences were first checked for adapters with Nesoni v0.13 (https://github.com/Victorian-Bioinformatics-Consortium/nesoni, last accessed February 11, 2020), trimmed at Q30 using prinseq v0.20.4 (http://prinseq.sourceforge.net, last accessed February 11, 2020), and checked for bacterial contamination with RAMBO-K v1.21 (Tausch et al. 2015) against the Refseq bacterial genomes to filter any remnants of contamination. These decontaminated sequences were assembled using SPAdes v3.6 (Bankevich et al. 2012) and corrected for local misassemblies and small INDELs using Pilon v1.13 (Walker et al. 2014). The genome annotation was performed using MAKER2 (Holt and Yandell 2011) in a two-step annotation workflow. Briefly, GeneMark-ES (Lomsadze et al. 2005) and CEGMA (Parra et al. 2007) were used to produce HMM profiles from the assembly. MAKER2 was then run a first time with these HMM profiles. The first annotation results were themselves converted into HMM profiles, and MAKER2 was run a second time using those updated HMM profiles. Annotation completeness was analyzed using CEGMA v2.5 (Parra et al. 2007) and BUSCO v2.0 (Simao et al. 2015). Protein sequences of the predicted genes were searched against the Swiss-Prot and TrEMBL databases (release 2019_07) using BlastP program (*e*-value $\leq 10^{-5}$). BLAST hits with query coverage <40% were filtered out. Functional annotation and classification of the annotated genes was performed using eggNOG mapper (Huerta-Cepas et al. 2016). Proteinortho v5.16 (Lechner et al. 2011) was used to find orthologous genes in the genomes with 30% identity, 50% coverage, and *e*-value $\leq 10^{-5}$.

Phylogenetic Analysis

The genomes of 40 nematodes were downloaded including 12 genomes from NCBI and 28 from WormBase version WS250 (Harris et al. 2010). At first, we identified "complete" BUSCO genes across the genomes of 41 nematodes using the Arthropoda database provided by BUSCO. Complete BUSCOs that were only present in <80% genomes were filtered out. Next, we aligned protein sequences of each BUSCO gene using MAFFT version 7.4 alignment program (Katoh et al. 2002). The resulting alignments were trimmed for spurious sequences or poorly aligned regions using trimAl version 1.4.1 (Capella-Gutierrez et al. 2009) with the "*-automated1*" set of parameters. A matrix of aligned sequences was created by concatenating all the trimmed alignments. To infer a species tree, the matrix was provided to IQ-tree software version 1.6.9 (Nguyen et al. 2015), run with ultrafast bootstrap replicates ($N = 1{,}000$) to determine the support for bipartitions and internode certainty.

Results and Discussion

Genome Assembly and Gene Annotation

In total 4 Gb of DNA sequence data were generated which is made up of 14,370,809 paired-end reads with a maximum read length of 301 bp and a minimum read length of 35 bp with a 44x raw sequence coverage. The GC content of the raw sequences was 34%. The genome was assembled into 8,974 scaffolds which is 89.88 Mb in length, with L50 of

Table 1
Summary of Assembly Statistics of the *Setaria digitata* Genome and Published Genomes of Several Nematodes That Cause Filariasis in Humans

Genomic Features	*Setaria digitata*[a]	*Loa loa*[b]	*Brugia malayi*[c]	*Wuchereria bancrofti*[d]
Sequencing technology	Illumina	454	WGS[e]	454
DNA (Mb)	89.8	91.3	93.6	88.4
DNA coding (Mb)	19.19	15.53	12.78	14.2
G+C content (%)	31.73	30.97	30.21	28.8
DNA scaffolds	8,974	5,764	24,286	5,105
N50 (bp)	24,961	174,388	41,308	56,670
L50 (bp)	882	130	226	351
Avg. N's per 100 kb	5.14	4210.81	7035.75	46.14
Protein-coding genes	20,568	15,440	11,460	11,068
Avg. gene length (bp)	3073	2989	2815	4307
Avg. exon per gene	8.6	6.8	7.15	9.2
Avg. exon length (bp)	109	164	158	141
Genes known in Swiss-Prot	6,009	6,157	5,263	5,855
Genes with signal peptides	711	—	—	—
Genes with TM helices	3,646	—	—	—
tRNA	174	124	97	—

[a]PRJEB13338 (present study).
[b]PRJNA37757.
[c]PRJNA27801.
[d]PRJNA275548.
[e]Whole-genome shotgun.

882 bp and an overall G + C content of 31.7%. The genome assembly covered over 87% of the total genome size, estimated using k-mer counts from the sequencing data. Additional genomic characteristics are shown in Table 1. The size and the GC content of the *S. digitata* genome were similar to *Loa loa* and *Brugia malayi* which cause filariasis in humans. Moreover, the number of predicted genes estimated among those filarial nematode genomes was comparable.

Gene prediction with MAKER2 annotation tool identified 20,568 protein-coding genes in the assembled genome (Table 1). The quality of the gene model data set was analyzed using the Swiss-Prot, KEGG, and Pfam databases. Putative function was assigned to 6,009 protein-coding genes based on the Swiss-Prot curated annotations, whereas 11,541 of the total genes were found in the TrEMBL database. In total, 8,560 (41.6%) of annotated genes have well-defined PFAM protein domains. Moreover, KEGG terms were assigned to 20% of the predicted genes. In total, the annotated regions comprise 21% of the genome with an average of 8.6 exons per gene and mean transcript length of 3,073 bp. We estimated the repeat content up to ~13.5% of the total genome. However, repeats annotation analysis could detect only 4.2% of the repetitive sequences, of which 2.9% were simple repeats. Furthermore, the quality of *S. digitata* genome annotation was assessed using CEGMA and BUSCO pipelines. Both CEGMA and BUSCO used conserved gene sets comprising 248 and 982 genes, respectively, to analyze the completeness of the genome. The analysis showed a high level of genome completeness (CEGMA 91.5%; BUSCO 85.5%) in the genome. Because BUSCO comprised a larger set of conserved genes compared with CEGMA, only the results of BUSCO analysis are shown in Table 2.

Table 2
BUSCO Statistics of the Completeness of the Genome Based on 982 Nematode-Conserved Genes

BUSCO	Genes Present	Percentage (%)
Complete BUSCOs (C)	839	85.5
Complete and single-copy BUSCOs (S)	805	82
Complete and duplicated BUSCOs (D)	34	3.5
Fragmented BUSCOs (F)	96	9.8
Missing BUSCOs (M)	47	4.7

Comparative Genomics and Phylogenetic Analysis

Comparative genomics of *S. digitata* with other filariasis causing parasitic nematodes *L. loa*, *B. malayi*, and *W. bancrofti*, revealed that 8,369 (40%) of *S. digitata* protein-coding genes were orthologous among the genomes, of which 4,643 genes were single-copy orthologs (supplementary table 1, Supplementary Material online). *Setaria digitata* shared 7,493 genes with *L. loa*, 7,070 with *W. bancrofti*, and 6,612 with *B. malayi*. In total, 5,087 genes were shared among the four nematodes. In addition, the functional classification of protein-coding genes classified 9,768 genes into

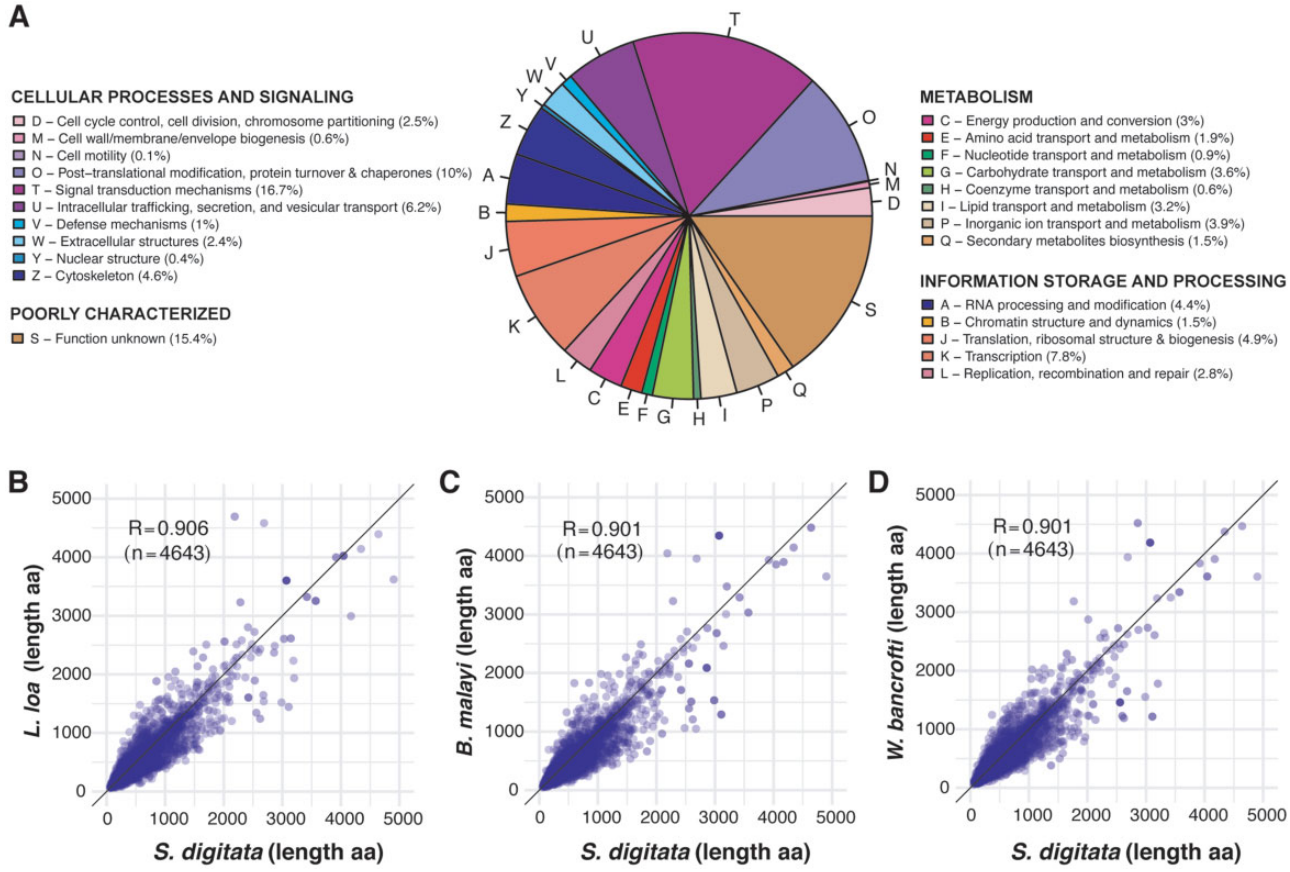

FIG. 1.—(*A*) Classification of protein-coding genes in COG functional groups. (*B–D*) Correlation of *Setaria digitata* proteins length with other nematode species. (*B*) *Setaria digitata* with *Loa loa*, (*C*) *S. digitata* with *Brugia malayi*, (*D*) *S. digitata* with *Wuchereria bancrofti*. The Pearson's correlation coefficients are calculated and shown on each plot.

different COGs (Cluster of Orthologs Groups) (fig. 1*A*, supplementary table 2, Supplementary Material online). Of the entire *S. digitata* gene set, 4,315 genes had KEGG orthologs linked to 375 KEGG biological pathways (supplementary table 3, Supplementary Material online).

A majority of filarial parasites, for instance, *W. bancrofti*, harbor an endobacterium, *Wolbachia*. Loss of *Wolbachia* in *Wolbachia*-dependent worm hampers growth and fertility of the host. Unlike many filarial species, both *S. digitata* and *L. loa* are devoid of *Wolbachia* endosymbiont (Voronin et al. 2015). However, the surface structure of *S. digitata* is similar to that of *W. bancrofti* (Madathiparambil et al. 2011). BLAST-based search of the assembled *S. digitata* genome against *Wolbachia* protein sequences did not reveal any large transfers of *Wolbachia* DNA.

Protein length correlation analysis of the single-copy orthologs showed a high level of correlation between the species (fig. 1*B–D*). *Setaria digitata* showed similar correlation ($R = 0.9$) with *L. loa*, *W. bancrofti*, and *B. malayi*. Strong correlations with the protein-coding genes of *W. bancrofti* and *L. loa* indicate coverage of protein-coding sequences identified in *S. digitata* is of considerably higher quality. Thus, genomic comparisons will greatly facilitate the identification of genes involved in development and potential drug targets.

To examine the evolution of filarial parasites in the context of other nematodes, we estimated a phylogeny across 41 nematode genomes available in NCBI and WormBase databases and *S. digitata* genome assembly created in this study (fig. 2). Maximum-Likelihood tree was generated based on 507 BUSCO genes present in at least 80% ($n \geq 32$) genomes. The tree was similar to 12S rDNA-based tree published earlier (Yatawara et al. 2007). Although the genus *Caenorhabditis* formed well-supported monophyletic group, *S. digitata* was grouped with the clade mainly consists of *Onchocerca* and *Brugia* species.

Conclusion

Here we present the genome sequence of *S. digitata*, a parasitic nematode found in the peritoneal cavity of cattle. Comparative genomic analysis of *S. digitata* with *W. bancrofti*, *L. loa*, and *B. malayi* genomes revealed similarity in several genomic features such as genome size, GC and

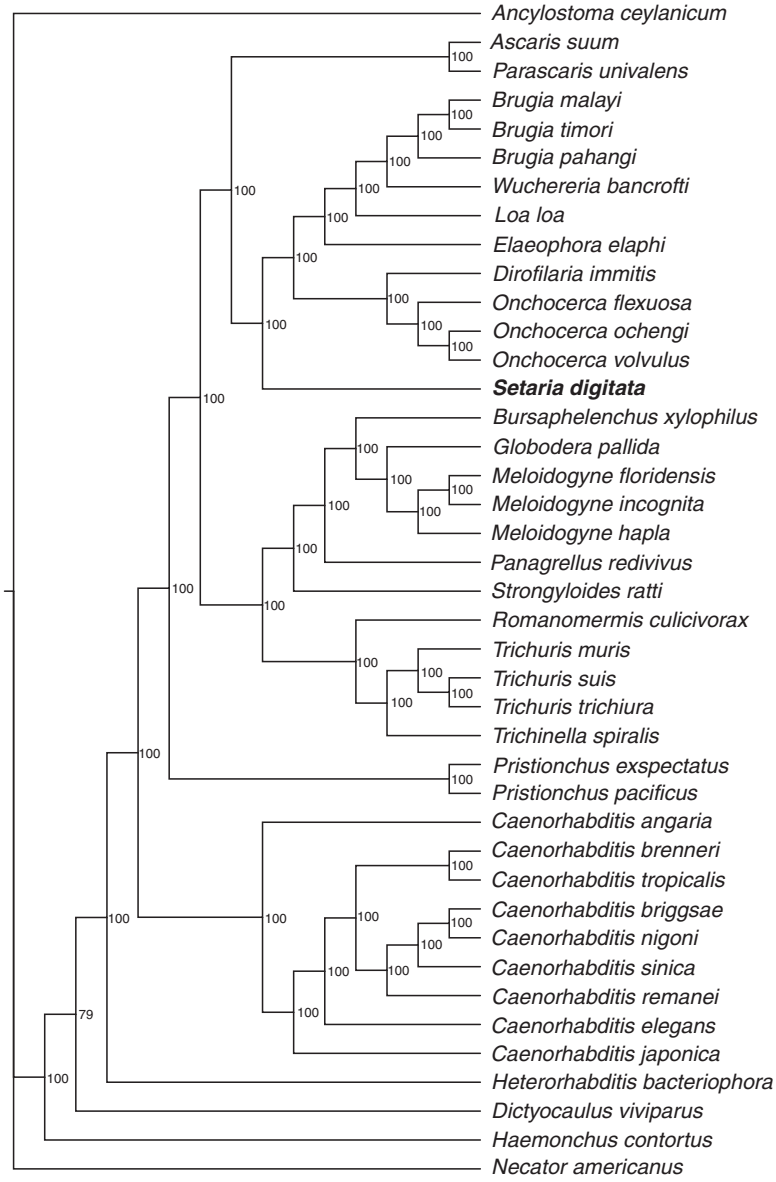

Fig. 2.—Phylogenetic analysis of the nematode genomes. Consensus tree between 41 nematode genomes, which includes the *Setaria digitata* assembly. Total 118,531 parsimony-informative sites out of 184,820 aligned positions were used to generate the tree under a single LG + G4 model. To the right of each node are bootstrap support values.

gene content. The genome of *S. digitata* presented here is comparable with the genomic architecture of other filarial parasites. Moreover, its morphological similarities to other filarial parasites such as *W. bancrofti*, and its better accessibility, support the use of *S. digitata* as a model organism for finding effective treatment for lymphatic filariasis. The availability of the genome of *S. digitata* will be a critical resource to not only elucidate the biology of filarial worms, but also benefit programs aimed at the treatment and elimination of these parasites. This draft genome will allow a more extensive manual curation of the annotation of the genome of *S. digitata*. Comparative analysis with the existent *W. bancrofti* draft genome is expected to lead to the identification of genes susceptible to new candidate drugs in both species that can be ultimately tested using *S. digitata* worms in laboratory conditions.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This study was funded by Swedish Research Links (Grant 2013-6757), awarded by the Swedish Research Council

(VR). This study was supported by the SLU Bioinformatics Infrastructure (SLUBI), Uppsala, Sweden. We thank Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden and Institute of Biochemistry, Molecular Biology and Biotechnology (IBMBB), University of Colombo, Sri Lanka for providing research facilities and other support for this study. We thank Hadrien Gourlé for his help with the assembly. We also acknowledge Oskar Karlsson Lindsjö for providing technical assistance with the sequencing

Literature Cited

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bazargani T, et al. 2008. Cerebrospinal nematodiasis of cattle, sheep and goats in Iran. Iran J Parasitol. 3(1):16–20.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973.

Decruse SW, Raj RK. 1990. Histological studies on female*Setaria digitata* (von Linstow 1906), a filaria of bovine. Proc Anim Sci. 99:103–112.

Dissanayake S, Ismail MM. 1980. Antigen of *Setaria digitata*: cross-reaction with surface antigens of *Wuchereria bancrofti* microfilariae and serum antibodies of *W. bancrofti*-infected subjects. Bull World Health Organ. 58(4):649–654.

Harris TW, et al. 2010. WormBase: a comprehensive resource for nematode research. Nucleic Acids Res. 38(Suppl 1):D463–467.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12(1):491.

Huerta-Cepas J, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44(D1):D286–293.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.

Kaur D, et al. 2015. Occurrence of *Setaria digitata* in a cow. J Parasit Dis. 39(3):477–478.

Lechner M, et al. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 12(1):124.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33(20):6494–6506.

Madathiparambil M, Raj NK, Linda J, Raj RK. 2011. Surface of the filarial parasite *Setaria digitata* and the structural changes on treatment with Triton X-100. Helminthologia 48(1):17–22.

Mohanty MC, Sahoo PK, Satapathy AK, Ravindran B. 2000. *Setaria digitata* infections in cattle: parasite load, microfilaraemia status and relationship to immune response. J Helminthol. 74(4):343–347.

Murugananthan A, Karunanayake EH, Tennekoon KH. 2010. Cloning and characterisation of alkali myosin light chain gene (MLC-3) of cattle filarial parasite *Setaria digitata*. IIOAB J. 1(4):1–10.

Nagaratnam N, Karunanayake EH, Tennekoon KH, Samarakoon SR, Mayan K. 2014. In silico characterization of a RNA binding protein of cattle filarial parasite *Setaria digitata*. Bioinformation 10(8):512–517.

Nakano H, et al. 2007. Morphological survey of bovine Setaria in the abdominal cavities of cattle in Aomori and Kumamoto Prefectures, Japan. J Vet Med Sci. 69(4):413–415.

Nayak A, Gayen P, Saini P, Mukherjee N, Babu SP. 2012. Molecular evidence of curcumin-induced apoptosis in the filarial worm *Setaria cervi*. Parasitol Res. 111(3):1173–1186.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9):1061–1067.

Perumal AN, Gunawardene YI, Dassanayake RS. 2016. *Setaria digitata* in advancing our knowledge of human lymphatic filariasis. J Helminthol. 90(2):129–138.

Rodrigo WW, Dassanayake RS, Weerasena SJ, Karunanayake EH. 2013. Heterologous expression of uncharacterized parasitic nematode-specific growth factor-like protein of *Setaria digitata* in *Pichia pastoris* expression systems. Trop Biomed. 30(2):181–192.

Rodrigo WW, Dassanayake RS, Weerasena SJ, Silva Gunawardene YI. 2014. Novel parasitic nematode-specific protein of bovine filarial parasite *Setaria digitata* displays conserved gene structure and ubiquitous expression. Trop Biomed. 31(3):514–524.

Shin J, et al. 2017. First blindness cases of horses infected with *Setaria digitata* (Nematoda: Filarioidea) in the Republic of Korea. Korean J Parasitol. 55(6):667–671.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212.

Small ST, et al. 2019. Human migration and the spread of the nematode Parasite *Wuchereria bancrofti*. Mol Biol Evol. 36(9):1931–1941.

Tausch SH, Renard BY, Nitsche A, Dabrowski PW. 2015. RAMBO-K: rapid and sensitive removal of background sequences from next generation sequencing data. PLoS One 10(9):e0137896.

Tung KC, Lai CH, Ooi HK, Yang CH, Wang JS. 2003. Cerebrospinal setariosis with *Setaria marshalli* and *Setaria digitata* infection in cattle. J Vet Med Sci. 65(9):977–983.

Voronin D, Abeykoon AM, Gunawardene YI, Dassanayake RS. 2015. Absence of Wolbachia endobacteria in Sri Lankan isolates of the nematode parasite of animals *Setaria digitata*. Vet Parasitol. 207(3–4):350–354.

Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

WHO. 1992. Lymphatic filariasis: the disease and its control. Fifth report of the WHO Expert Committee on Filariasis. World Health Organ Tech Rep Ser. 821:1–71.

Associate editor: Sandra Baldauf