Biotechnology Advances xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

# Cracking the nodule worm code advances knowledge of parasite biology and biotechnology to tackle major diseases of livestock

Rahul Tyagi [a], Anja Joachim [b], Bärbel Ruttkowski [b], Bruce A. Rosa [a], John C. Martin [a], Kymberlie Hallsworth-Pepin [a], Xu Zhang [a], Philip Ozersky [a], Richard K. Wilson [a], Shoba Ranganathan [c], Paul W. Sternberg [d], Robin B. Gasser [e,*], Makedonka Mitreva [a,f,**]

[a] The Genome Institute, Washington University in St. Louis, MO 63108, USA
[b] Institute of Parasitology, Department of Pathobiology, University of Veterinary Medicine Vienna, Veterinärplatz 1, A-1210 Vienna, Austria
[c] Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales, Australia
[d] HHMI, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA
[e] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia
[f] Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA

ARTICLE INFO

*Article history:*
Received 13 April 2015
Accepted 23 May 2015
Available online xxxx

*Keywords:*
Biotechnology
Genomics
Transcriptomics
Bioinformatics
Nodule worm disease
Livestock

ABSTRACT

Many infectious diseases caused by eukaryotic pathogens have a devastating, long-term impact on animal health and welfare. Hundreds of millions of animals are affected by parasitic nematodes of the order Strongylida. Unlocking the molecular biology of representatives of this order, and understanding nematode–host interactions, drug resistance and disease using advanced technologies could lead to entirely new ways of controlling the diseases that they cause. *Oesophagostomum dentatum* (nodule worm; superfamily Strongyloidea) is an economically important strongylid nematode parasite of swine worldwide. The present article reports recent advances made in biology and animal biotechnology through the draft genome and developmental transcriptome of *O. dentatum*, in order to support biological research of this and related parasitic nematodes as well as the search for new and improved interventions. This first genome of any member of the Strongyloidea is 443 Mb in size and predicted to encode 25,291 protein-coding genes. Here, we review the dynamics of transcription throughout the life cycle of *O. dentatum*, describe double-stranded RNA interference (RNAi) machinery and infer molecules involved in development and reproduction, and in inducing or modulating immune responses or disease. The secretome predicted for *O. dentatum* is particularly rich in peptidases linked to interactions with host tissues and/or feeding activity, and a diverse array of molecules likely involved in immune responses. This research progress provides an important resource for future comparative genomic and molecular biological investigations as well as for biotechnological research toward new anthelmintics, vaccines and diagnostic tests.



## 1. Introduction

*Oesophagostomum dentatum* (nodule worm) is an economically important parasite of swine; this dioecious nematode belongs to the large order of the Strongylida (strongylids) that infect humans and animals worldwide (Anderson, 2000; Taylor et al., 2007). For instance, more than 1.3 billion people are infected with strongylids, such as *Necator americanus* and/or *Ancylostoma duodenale* (hookworms), which feed on blood in the small intestine (Hotez et al., 2013), causing adverse effects on human health, particularly in children. Other strongylids infect livestock and cause substantial production losses due to subclinical infections and clinical disease (Cantacessi et al., 2012; Lichtenfels et al., 1997), with billions of dollars spent annually on treatments to control these worms. In addition to their socioeconomic impact, various strongylid nematodes have developed resistance against the main drug classes commonly used to treat the diseases that they cause (Gilleard, 2006). Therefore, there is a need to work toward new treatment or control methods. This quest should be facilitated by acquiring a deep understanding of the molecular biology and biochemistry of key representatives.

*O. dentatum* is transmitted orally to the host and has a complex 3-week life cycle (Christensen et al., 1995; Spindler, 1933) (Fig. 1): eggs are excreted in host feces; the first-stage larva (L1) develops inside the egg to then hatch (within 1 day) and molt through to the second- (L2) and third-stage (L3) larvae within a week; the infective L3s are then ingested by the host, exsheath and, after a histotrophic phase, develop through the fourth-stage larvae (L4s) to dioecious adults, which feed on nutrients in the large intestine. Because of its short life

* Corresponding author.
** Correspondence to: M. Mitreva, The Genome Institute, Washington University in St. Louis, MO 63108, USA.
*E-mail addresses:* robinbg@unimelb.edu.au (R.B. Gasser), mmitreva@genome.wustl.edu (M. Mitreva).

http://dx.doi.org/10.1016/j.biotechadv.2015.05.004
0734-9750/

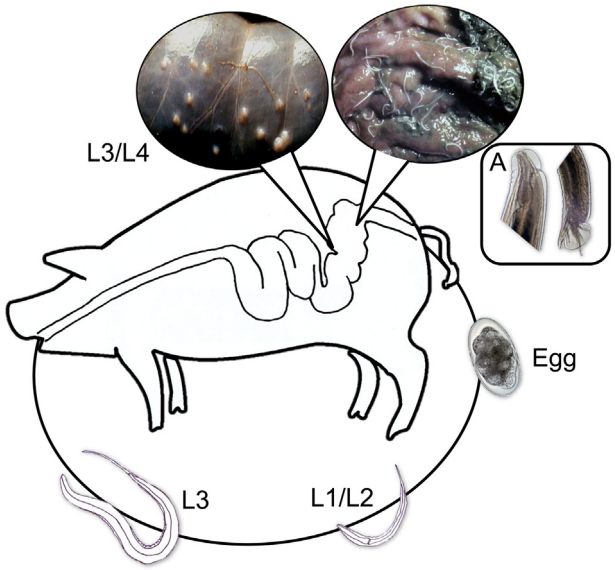


**Fig. 1.** Development of *Oesophagostomum dentatum*. In a 3-week life cycle of the parasite, eggs are excreted in host feces; the first-stage larva (L1) develops inside the egg to hatch and molt through to the second- (L2) and third-stage (L3) larval stages within a week; the infective L3s are then ingested by the pig host, exsheath and, after a short tissue phase, develop through the fourth-stage larvae (L4) to dioecious adults (A) which both feed on mucus and the contents of the large intestines.

cycle and ability to develop in vitro for weeks through several molts (Daugschies and Watzel, 1999), *O. dentatum* is a useful model system for profound investigations of fundamental biological processes in nematodes. What has been missing, however, is basic information on the genome and transcriptomes to underpin such explorations.

Recent advances in the sequencing of the draft genomes and transcriptomes of selected, blood-feeding strongylid nematodes *Haemonchus contortus* (barber's pole worm of sheep) and *N. americanus* (Schwarz et al., 2013; Tang et al., 2014) now provide a sound basis to tackle related nematodes that do not feed on blood. Here, we report on recent advances made in biology and biotechnology through the draft genome and developmentally-staged transcriptome to substantially enhance our understanding of this pathogen at the molecular level, across all defined life cycle stages, and its relationship with the porcine host. This genome not only delivers an important resource to the scientific community for a wide spectrum of genomic, systematic, biological and epidemiological studies, it also provides a solid foundation for the development of new interventions (drugs, vaccines and diagnostic tests) against *O. dentatum* and related nematodes of the superfamily Strongyloidea.

## 2. Genome characteristics and protein-encoding gene sets

The nuclear genome of *O. dentatum* (Od-Hann strain) was sequenced, assembled and annotated (Appendix A). The final draft assembly was 443 Mb, half of which was represented by supercontigs of >34.6 kb. This genome is among the largest of any nematode studied to date (Table 1), and the assembly represents most (90%) of the genome (Parra et al., 2007, 2009). The GC content is 41.3%, and the estimated repeat content is 30.9%, equating to 136.7 Mb of the genome. In total, 1591 repeat families were predicted and annotated (Table 1); 350 transposons, including 137 DNA transposons, 112 LTR retrotransposons and 89 non-LTR retrotransposons were identified among these repeat families. The protein-encoding genes predicted ($n = 25{,}291$) represent 12.4% of the genome and have an average density of 57 genes per Mb. The GC content of coding sequences is 47.2%. Functional annotation of predicted proteins by sequence comparisons identified 4540 unique domains (IPR) and 1354 gene ontology (GO) terms for 62% and 48% of the *O. dentatum* genes, respectively; KEGG-based annotations were assigned to 56% of the proteins predicted for *O. dentatum*, which represented 4171 unique KEGG orthology groups (KOs). Comparison of the gene set of *O. dentatum* with those of three other nematodes from the same taxonomic order (Blaxter et al., 1998) identified 15,076 orthologous groups (OGs) (Fig. 2); 11,213 of these OGs contained at least one *O. dentatum* gene (66.6% of all protein-coding genes), 2699 OGs included genes from at least one other nematode but none from *Caenorhabditis elegans*, and 1044 OGs (representing 2972 genes) were unique to *O. dentatum*.

Some common drug target entities, including kinases, phosphatases, GTPases, G protein-coupled receptors (GPCRs) and transporters, were annotated. In total, 327 kinases and 445 phosphatases were encoded in the genome of *O. dentatum* (Table 1). All of the major classes of kinases were identified, with tyrosine (TK; $n = 76$), CAMK (51), CMGC (49) and casein kinases (CK1; 45) being abundantly represented (67.6%). The phosphatases annotated include mainly protein tyrosine ($n = 77$), serine/threonine (63), dual specificity (56) and histidine (41) phosphatases. In addition, 159 GTPases including 91 small GTPases within the families Rab ($n = 68$), Rho (42), Arf (9) and Ran (6) were predicted. A small number ($n = 12$) of large GTPases, including dynamin, GBP and mitofusin (Table 1), were also identified. Examples of salient small GTPase homologs are *rab-5*, *ran-1*, *sar-1*, *mig-2* and *rheb-1*, whose

**Table 1**
Features of the genome of *Oesophagostomum dentatum* and three other nematode species (*Haemonchus contortus*, *Necator americanus* and *Caenorhabditis elegans*) included for comparative purposes.

| | *O. dentatum* | *H. contortus* | *N. americanus* | *C. elegans* |
|---|---|---|---|---|
| Estimated genome size (Mega bases) | 443 | 320 | 244 | 100 |
| Assembly statistics | | | | |
| Total number of supercontigs (≥1 kb) | 64,258 | 14,419 | 11,713 | 6 |
| Total number of base pairs (bp) in supercontigs | 443,038,381 | 319,640,208 | 244,009,025 | 100,272,607 |
| Number of N50 supercontigs | 3096 | 1,684 | 283 | 3 |
| N50 supercontig length (bp) | 26,460 | 56,328 | 213,095 | 17.5Mb |
| Number of N90 supercontigs | 31,598 | 6085 | 1336 | 6 |
| N90 supercontig length (bp) | 2153 | 13,105 | 29,214 | |
| GC content of whole genome (%) | 33% | 42.40% | 40.20% | 35.40% |
| Repetitive sequences (%) | 30.86% | 13.40% | 24% | 21% |
| Protein-coding loci | | | | |
| Total number of protein coding genes | 25,291 | 23,610 | 19,151 | 20,517 |
| Mean gene loci footprint (bp) | 2171 | 6167 | 4289 | 3035 |
| Mean number of exons per gene | 5.3 | 7 | 6.4 | 6.1 |
| Mean exon size (bp) | 122 | 139 | 125 | 203 |
| Mean intron size (bp) | 352 | 832 | 642 | 320 |

N50: number-50% of all nucleotides in the assembly are in 3096 supercontigs, length-50% of the genome is in supercontigs with a minimum length of 26 kb; N90: number-90% of all nucleotides in the assembly are within 31,598 supercontigs, length-90% of the genome is in supercontigs with a minimum length of 2153.
(considering space on either strand)/(considering same strand space).

Please cite this article as: Tyagi, R., et al., Cracking the nodule worm code advances knowledge of parasite biology and biotechnology to tackle major diseases of livestock, Biotechnol Adv (2015), http://dx.doi.org/10.1016/j.biotechadv.2015.05.004