2009

Integrating genomics and phylogenetics in understanding the history of Trichinella species

Dante Zarlenga

B. Rosenthal

E. Hoberg

M. Mitreva

Veterinary Parasitology 159 (2009) 210–213

Contents lists available at ScienceDirect

Veterinary Parasitology

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

Integrating genomics and phylogenetics in understanding the history of *Trichinella* species

D.S. Zarlenga [a,b,*], B. Rosenthal [b], E. Hoberg [b], M. Mitreva [c]

[a] Bovine Functional Genomics Lab, USDA, ARS, ANRI, Beltsville, MD 20705, USA
[b] Animal Parasitic Diseases Lab, USDA, ARS, ANRI, Beltsville, MD 20705, USA
[c] Genome Sequencing Center, Washington University School of Medicine, St. Louis, MO 63108, USA

ARTICLE INFO

Keywords:
Trichinella
Phylogeny
Genome sequencing
Nematode

ABSTRACT

In 2004, funding was received by Washington University's Genome Sequencing Center through NHGRI, to completely sequence several nematode genomes as part of a holistic effort to advance our understanding of the human genome and evolution within the Metazoa. *Trichinella spiralis* was among this group of worms because of its strategic location at the base of the phylum Nematoda, and the belief that extant species represented an ancient divergent event that occurred as early as the Paleozoic. At the same time, a concerted effort was put forth to solidify the phylogeny of extant species of *Trichinella* based upon molecular analyses of a multi-gene system to understand the history of the genus and thereby enhance utilization of the forthcoming sequence data. Since the inception of this research, several findings have emerged: (1) the size of *T. spiralis* genome estimated by flow cytometry (71.3 Mb) is substantially smaller than originally predicted (270 Mb); (2) to date, a subset of the total of 3,534,683 sequences have been assembled into a 59.3 Mb unique sequence; (3) 19% of the assembled sequence is comprised of repetitive elements; and (4) sequence data are predicated upon extant *T. spiralis* which probably diverged as little as 20 million years ago. Thus, the utility of the *T. spiralis* genome as representative of an archaic species must be tempered with the knowledge that encapsulated and non-encapsulated clades probably separated during the mid-Miocene as temperate ecosystems changed.

1. Introduction

Although eating raw or undercooked pork infected with the larvae of *Trichinella spiralis* accounts for most human trichinellosis, 11 distinct forms have been recognized in mammals, crocodilians, and birds (Pozio and Zarlenga, 2005). These genotypes can be morphologically delineated into two clades by the presence or absence of a cyst encapsulating both the muscle cell and the intracellular, infectious stage. A better appreciation of the evolutionary history of this genus would substantially advance the comparative value of the complete genome sequence of *T. spiralis* as a representative model of a basal nematode. In addition, understanding the phylogeny and biogeography of extant *Trichinella* may elucidate past and present determinants to zoonotic risk and provide an opportunity to understand how complex host–parasite systems diversify in ecological and evolutionary time. To this end, the first robust, multi-locus phylogenetic analysis of all ecologically and genetically recognized species and genotypes of *Trichinella* was generated and evaluated it in relation to the history of their vertebrate hosts (Zarlenga et al., 2006). Historical/phylogenetic analyses exploring the coevolutionary relationships of hosts and parasites

* Corresponding author at: U.S. Department of Agriculture, ARS, Bovine Functional Genomics and Animal Parasitic Diseases Laboratories, Bldg. 1180, BARC-East, Beltsville, MD 20705, United States.
Tel.: +1 301 504 8754; fax: +1 301 504 8979.
E-mail address: Dante.Zarlenga@ars.usda.gov (D.S. Zarlenga).

0304-4017 Published by Elsevier B.V. Open access under CC BY-NC-ND license.
doi:10.1016/j.vetpar.2008.10.061

refuted the hypothesis for origins and diversification of extant species of *Trichinella* that coincided with initial radiation of vertebrates during the Paleozoic. Although the lineage for *Trichinella* is archaic, extant species were shown to have distributions consistent with associations primarily among mammalian hosts and patterns of diversification that have extended over the past 20 million years in relatively shallow evolutionary time (Zarlenga et al., 2006).

The value of *Caenorhabditis elegans* as a model for biomedical research is well documented. The decoding of its genome has been enhanced by draft sequence data from other *Caenorhabditis* species to allow better alignment, gene interpretation, promoter analysis, identification of non-coding RNA and other functional features; however, genome sequence comparison among congeners is woefully inadequate for extrapolation across disparate genera, families or higher taxonomic groups. Coupled with the knowledge that gene loss was strategically important in the evolution of *C. elegans* (Parkinson et al., 2004) and that parasitism among nematode groups in vertebrates evolved independently (Blaxter et al., 1998), the study of parasitism is unlikely to be advanced by exclusive attention to models based on free-living nematode models. Thus, as a component of a broader program to elucidate the disjunction between parasitism and ancient evolution within this Phylum, to better understand the biology and diversity of nematodes, and to evaluate the relationship between Nematoda and Metazoa in general, the genome of *T. spiralis* is now being sequenced in its entirety.

2. Materials and methods

The *T. spiralis* genome size was estimated by propidium iodide (PI) based flow cytometry of isolated nuclei. Adult parasites were washed in cold Galbraith buffer (GB: 0.5 M $MgCl_2$, 0.3 M sodium citrate, 0.02 M MOPS, 0.1% Triton X-100, 1 μg/ml boiled ribonuclease A, pH 7.2), chopped (50×) with a razor blade in 30–50 μl cold GB buffer, then transferred into a 1.5 ml Dounce homogenizer using 1 ml of buffer. Similarly prepared specimens of *C. elegans* and the head of a *Drosophila melanogaster* (Iso-1) and/or *D. virilis* female were added as standards. The mixture was ground (15 strokes of the "A" pestle at a rate of 3 strokes per two seconds), passed through 20 μm nylon filter and adjusted to 1 ml with cold GB. Propidium iodide (1 mg/ml H_2O) was added to a final concentration of 75 μM, then stored in the dark for 2–8 h. Samples were run in a Beckman-Coulter Epics Elite Cytometer using 25 mW of 488 nm (blue) excitation. Fluorescence was measured across a 610 nm pass filter. Counting was activated by PI fluorescence. Only those nuclei with the lowest scatter level (intact, untagged nuclei) were counted. Five replicates were performed.

The *T. spiralis* genome sequencing project is still in progress at the Washington University's Genome Sequencing Center. The first version of the sequence assembly (http://genome.wustl.edu/genome.cgi?GENOME=Trichinella%20spiralis&SECTION=assemblies) was generated using the PCAP package (Huang et al., 2006) and used paired-end sequences from plasmid, fosmid and bacterial artificial chromosome (BAC)-end clones. A BAC-based physical map is being generated along with primer-directed sequence improvement on supercontigs and regions of low sequence coverage to close gaps and improve the assembly. Sequence data for *T. pseudospiralis* is currently being generated using 454 technologies (Margulies et al., 2005).

3. Results and discussion

The size of the *T. spiralis* genome was estimated by flow cytometry to be 1C = 71.3 + /1 1.2 Mb. *T. spiralis* ran below the standards for *C. elegans* and *Drosophila* in every replicate (*C. elegans* 1C = 100 Mb, *D. melanogaster* 1C = 175 Mb, and *D. virilis* 1C = 333.2 Mb). This size is substantially smaller than previously estimated (26 pg; 1 pg = 978 Mb); however, biochemical techniques are less accurate for larger genomes and likely account for the discrepancy. This estimate is the average from male and female worms. Assuming both sexes were present in equal proportions, the male would have a predictive size of 68 Mbp (1C = 67.7), while the female would be 75 Mbp (1C = 74.9). Two genome size peaks were observed in the flow cytometry data; however, these were insufficiently resolved (<6% difference).

The first assembly contained a total of 3,534,683 reads (3,347,411 genome survey sequences [GSSs], 149,088 fosmid ends, 38,184 BAC ends). Of these 96% were assembled in 9282 contigs and subsequently into 6262 supercontigs. The total length of the contigs was 59.3 Mb (34× coverage). This size is in reasonable agreement with that predicted by flow cytometry (71.3 Mb). Less complete sequence data on *T. pseudospiralis* have also been generated for comparative purposes (Table 1). Of the total 25,071,070 bp of the *T. pseudospiralis assembly*, 15,792,837 bp mapped back to the *T. spiralis* assembly (63%; cut-off 60 bits and 60% identity); however, only 10.9 Mb has been sequenced thus far. Additional coverage of the *T. pseudospiralis* genome is in progress.

The repeat content in the *T. spiralis* assembly v.1 was evaluated by masking simple repeats, low-complexity repeats and repeats identified by generating a custom library of repeat sequences. A custom library constructed using RECON (Bao and Eddy, 2002) and default parameters was screened for non-coding RNA and protein-coding genes using Rfam and non-redundant GenBank (built April 2007), respectively. RECON is designed to identify repeat

Table 1
Genome sequence data from *Trichinella pseudospiralis*.

T. pseudospiralis[a]	Reads	Assembly[b]
Total sequences	424,371	73,526
Total length (bp)	109,882,781	25,071,770
Min length (bp)	37	37
Max length (bp)	455	13,881
Mean read (bp)	259	341
Mean length S.D.	43	281
Median length (bp)	266	258

[a] These data were generated by single run on the 454 Life Sciences FLX sequencer.
[b] Newbler assembly (contigs & singletons).

Table 2
Sequence comparisons among nematode genomes.

	C. elegans	*B. malayi*	*T. spiralis*
Assembly	100	70.8	59.3
Overall %GC	35.4	30.5	33.9
% Repeat	16.5	15	19.1
		# hits	% id
Ce[a] vs Bm[b]		10,216	47.5
Ce vs Ts		9,288	41.7

[a] WormPep158 was used (23,162 genes).
[b] TIGR assembly (downloaded on 070606, 8180 contigs).

element families in an assembly but does not provide information on the identity of elements. Results yielded 376 repeat families. RepeatMasker (http://repeatmasker.org) was used to estimate the total repeat content at 19.1% out of which 1.0% are simple repeats, 2.5% are low complexity repeats, and 15.6% are repeats identified by the custom library. Characterization of the repeats is in progress.

Comparisons were drawn among datasets available for *C. elegans* (downloaded from WormBase, April 2007), *Brugia malayi* (Ghedin et al., 2007), and *T. spiralis* along with newly generated sequence information from *Trichinella pseudospiralis* (Table 2). Reasonable congruence was observed in the overall % repeat sequences, %GC content and sequence identity among the groups (41–47%). The low level of sequence identity among phylogenetically disparate groups of nematodes is consistent with conservation being relegated to regions found among broader taxa rather than limited only to members of this phylum.

Although divergence from a common ancestor for trichurids and trichinellids extends to the late Paleozoic or early Mesozoic, the concept for a deep history for contemporary *Trichinella* has been effectively refuted (Fig. 1) (Zarlenga et al., 2006). Relatively shallow divergence times for species within the encapsulated and non-encapsulated lineages and for the respective subclades suggest divergence likely occurred within a timeframe of 15–30 MYBP. In addition, a history of association with Eutheria with secondary colonization of other tetrapod groups is apparent where transmission was linked to guild associations among mammals. We contend, therefore, that contemporary diversity within the genus is limited to the Tertiary and the age of eutherian mammals, and is linked to multiple events of biotic expansion from Eurasia to Africa after the Miocene and from Eurasia into the Nearctic during the Pliocene and Quaternary.

Divergence estimates for the separation of organisms belonging to *Trichinella* and *Trichuris* have recently been

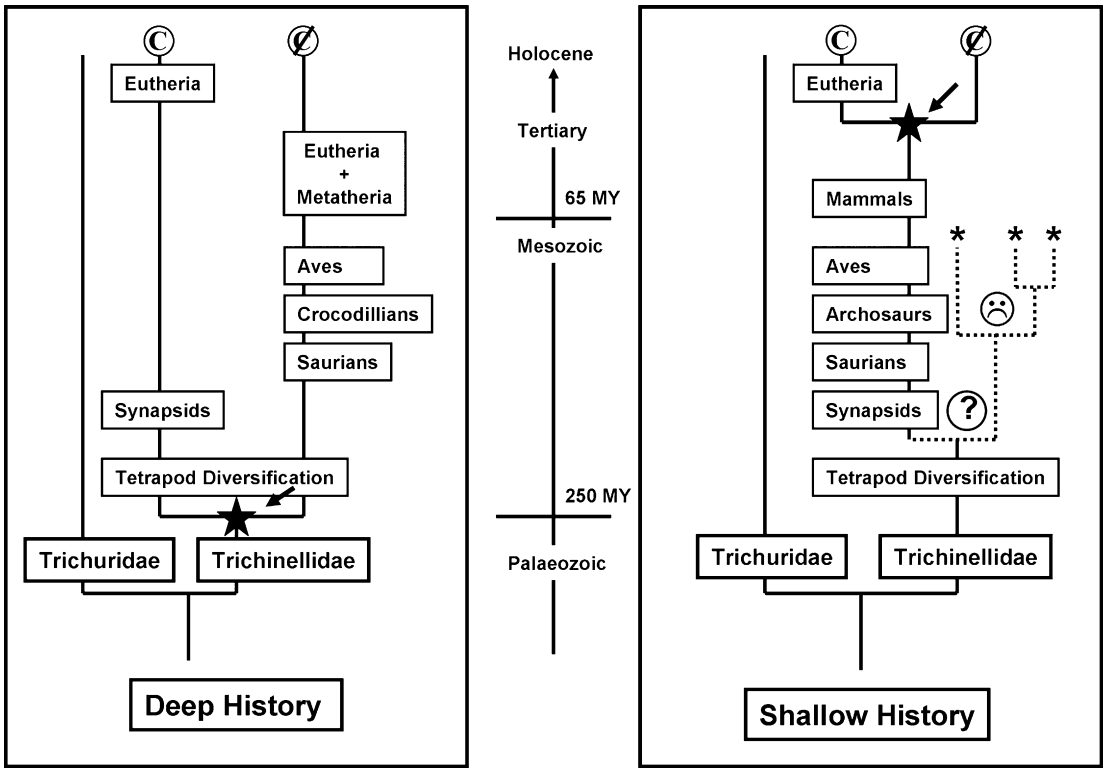

Fig. 1. Two general hypotheses have been advanced for the diversification of *Trichinella*. Phylogenetic data are consistent with extant species having associations primarily with eutherian hosts, and demonstrating patterns of diversification among encapsulated © and non-encapsulated ∅ clades that extended over the past 15–20 million years in relatively shallow evolutionary time. Contemporary diversity raises questions (?) regarding ancestral associations.

estimated at 250–300 MYBP using a multi-gene system which includes the small subunit ribosomal RNA gene (Zarlenga et al., 2006). This contrasts greatly with the estimate of 600–700 MYBP for the divergence of lineages leading to *C. elegans* and *T. spiralis* based upon globin and cytochrome C amino acid sequence data (Mitreva and Jasmer, 2006; Stein et al., 2003), and with the divergence of nematodes from arthropods that has been postulated at 800–1000 MYBP (Blaxter, 1998; Wang et al., 1999). The postulated time of divergence for the trichinellids and the trichurids, however, corroborates theories of the origins of the metazoan phyla that coincide with paleontological estimates proposed by Ayala et al. (1998) which assumed a faster-evolutionary rate for invertebrates. Nonetheless, the age discrepancy between the divergence of basal nematodes and the apparent young age of all extant *Trichinella* species raises questions as to the utility of comparative genomics to provide a picture of the evolutionary landscape across the extensive diversity represented in the phylum. Care must therefore be taken in extrapolating current sequence data as representative of an archaic genus.

Conflict of interest statement

No financial or personal relationships are maintained with other people or organizations that could inappropriately influence or bias this paper.

Acknowledgements

The National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH) is providing funding for the sequence characterization of the *Trichinella spiralis* genome.

References

Ayala, F.J., Rzhetsky, A., Ayala, F.J., 1998. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proc. Natl. Acad. Sci. U.S.A. 95, 606–611.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12, 1269–1276.

Blaxter, M., 1998. *Caenorhabditis elegans* is a nematode. Science 282, 2041–2046.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., et al., 1998. A molecular evolutionary framework for the phylum Nematoda. Nature 392, 71–75.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D., et al., 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317, 1756–1760.

Huang, X., Yang, S.-P., Chinwalla, A.T., Hillier, L.W., Minx, P., Mardis, E.R., Wilson, R.K., 2006. Application of a superword array in genome assembly. Nucleic Acids Res. 34, 201–205.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Mitreva, M., Jasmer, D.P., 2006. Biology and genome of *Trichinella spiralis*. In: The *C. elegans* Research Community (Ed.), WormBook, doi/10.1895/wormbook.1.124.1 <http://www.wormbook.org>.

Parkinson, J., Mitreva, M., Whitton, C., Thomson, M., Daub, J., Martin, J., Hall, N., Barrell, B., Waterston, R.H., McCarter, J.P., Blaxter, M., 2004. A transcriptomic analysis of the phylum Nematoda. Nat. Genet. 36, 1259–1267.

Pozio, E., Zarlenga, D.S., 2005. Recent advances on the taxonomy, systematics and epidemiology of *Trichinella*. Int. J. Parasitol. 35, 1191–1204.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al., 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1, E45.

Wang, D.Y., Kumar, S., Hedges, S.B., 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. Biol. Sci. 266, 163–171.

Zarlenga, D.S., Rosenthal, B.M., La Rosa, G., Pozio, E., Hoberg, E.P., 2006. Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. Proc. Natl. Acad. Sci. U.S.A. 103, 7354–7359.