

RESEARCH

Open Access



# Mitochondrial and ribosomal markers in the identification of nematodes of clinical and veterinary importance

María José Mejías-Alpizar<sup>1†</sup>, Catalina Porras-Silesky<sup>1†</sup>, Esteban José Rodríguez<sup>2</sup>, Joban Quesada<sup>1</sup>, María Paula Alfaro-Segura<sup>1</sup>, Joby Robleto-Quesada<sup>2</sup>, Ricardo Gutiérrez<sup>3,4</sup> and Alicia Rojas<sup>1,2\*</sup>

## Abstract

**Background** Nematodes of the Ascarididae, Ancylostomatidae and Onchocercidae families are parasites of human and veterinary importance causing infections with high prevalence worldwide. Molecular tools have significantly improved the diagnosis of these helminthiases, but the selection of genetic markers for PCR or metabarcoding purposes is often challenging because of the resolution these may show.

**Methods** Nuclear 18S rRNA, internal transcribed spacers 1 (ITS-1) and 2 (ITS-2), mitochondrial gene cytochrome oxidase 1 (*cox1*) and mitochondrial rRNA genes 12S and 16S loci were studied for 30 species of the mentioned families. Accordingly, their phylogenetic interspecies resolution, pairwise nucleotide p-distances and sequence availability in GenBank were analyzed.

**Results** The 18S rRNA showed the least interspecies resolution since separate species of the *Ascaris*, *Mansonella*, *Toxocara* or *Ancylostoma* genus were intermixed in phylogenetic trees as opposed to the ITS-1, ITS-2, *cox1*, 12S and 16S loci. Moreover, pairwise nucleotide p-distances were significantly different in the 18S compared to the other loci, with an average of  $99.1 \pm 0.1\%$ ,  $99.8 \pm 0.1\%$  and  $98.8 \pm 0.9\%$  for the Ascarididae, Ancylostomatidae and Onchocercidae families, respectively. However, ITS-1 and ITS-2 average pairwise nucleotide p-distances in the three families ranged from 72.7% to 87.3%, and the *cox1*, 12S and 16S ranged from 86.4% to 90.4%. Additionally, 2491 *cox1* sequences were retrieved from the 30 analyzed species in GenBank, whereas 212, 1082, 994, 428 and 143 sequences could be obtained from the 18S, ITS-1, ITS-2, 12S and 16S markers, respectively.

**Conclusions** The use of the *cox1* gene is recommended because of the high interspecies resolution and the large number of sequences available in databases. Importantly, confirmation of the identity of an unknown specimen should always be complemented with the careful morphological examination of worms and the analysis of other markers used for specific parasitic groups.

**Keywords** Molecular diagnosis, Phylogenetic analysis, Mitochondrial markers, Ribosomal markers, Ascarididae, Ancylostomatidae, Onchocercidae

<sup>†</sup>María José Mejías-Alpizar and Catalina Porras-Silesky contributed equally to this work.

\*Correspondence:

Alicia Rojas

anaalicia.rojas@ucr.ac.cr

Full list of author information is available at the end of the article



## Background

Nematodes of the families Ascarididae, Ancylostomatidae and Onchocercidae are parasites of human and veterinary importance causing infections with high prevalence worldwide [1–3]. Many of the infections produced in humans are cataloged as tropical neglected diseases, associated to areas with low socioeconomic index and poor access to diagnostic tools and treatment [4, 5].

Members of the families Ascarididae, Ancylostomatidae and Onchocercidae may cause clinical manifestations in virtually all known mammal species, whom get infected by the intake of contaminated water or food, contact with soil or an intermediate host [6]. For instance, the human ascarid *Ascaris lumbricoides* is the most common human nematode reported worldwide, with up to 1.2 billion estimated people infected [7]. In the veterinary field, *Parascaris equorum* and *Baylisascaris procyonis* produce intestinal infections in equines and raccoons [8], respectively, whereas *Toxocara canis* generates infections in > 100 million dogs worldwide [9] and can produce a zoonotic disease in humans, known as visceral, ocular or cerebral larva migrans with an estimated global human seroprevalence of 19% [9, 10]. Furthermore, blood-sucking hookworms like *Necator americanus* and *Ancylostoma* spp. (Ancylostomatidae family) are responsible for 740 million people infected worldwide causing intestinal bleeding and chronic ferropenic anemia [4]. Finally, filarioids such as *Wuchereria bancrofti*, *Brugia malayi*, *Onchocerca* spp. and *Dirofilaria* spp. are vector-borne heteroxenous parasites of the family Onchocercidae with worldwide distribution that affect thousands of humans, companion animals, livestock and wildlife [5].

Sensitive diagnostic tools are essential to improve clinical detection and treatment in animal and human hosts. However, traditional parasitological methods of diagnosis usually show low sensitivity for detecting and identifying species [3]. For instance, although direct microscopic observation of stool samples and concentration methods like Kato-Katz, Sheather or Ritchie detect ascarid and hookworm eggs and larvae using unsophisticated equipment and inexpensive reagents [11], these methods are negatively affected by low parasitic loads, intermittent excretion of eggs and uneven distribution of eggs in stool samples, significantly affecting their sensitivity [12]. Similarly, the diagnosis of filariasis is based predominantly on the microscopic detection of circulating microfilariae in blood, either by direct blood smears or Knott's modified test, or the recognition of the worm antigens using immunochromatographic tests. However, false-negative results are common in low parasitemia, single-sex or mixed infections. To alleviate these limitations of classical methods in parasitology, molecular methods with

higher sensitivity and specificity have been designed [13, 14].

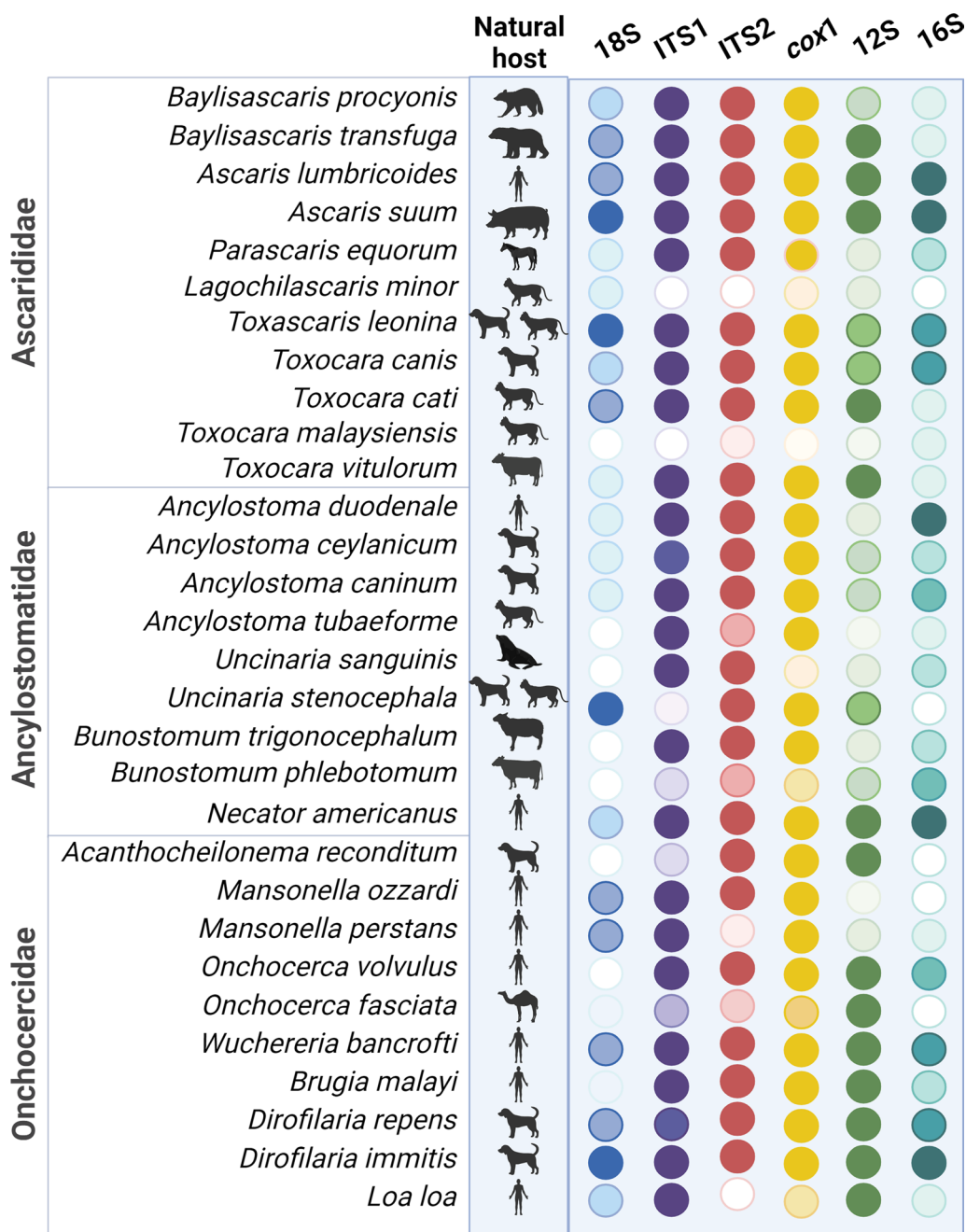
Molecular tools have improved the diagnosis of infections and facilitated taxa differentiation and phylogenetic analysis of worms of the Ascarididae, Ancylostomatidae and Onchocercidae families [14, 15]. Mitochondrial genes such as cytochrome c oxidase subunit 1 (*cox1*), cytochrome b (*cytb*), and the NADH dehydrogenase subunit 1 (*nad1*), and mitochondrial rRNA 12S and 16S are widely used to determine lower taxonomic levels (i.e. identification to genus and species-level) and resolve species-level phylogenies because of their high degree of sequence variation [16]. Similarly, internal transcribed spacers (ITS) 1 and 2 of the ribosomal nuclear DNA also show a high degree of sequence variation, promoting their use as markers for species differentiation [16]. Contrarily, the 18S nuclear rRNA gene is highly conserved within species and has low inter-species variability but has been employed for characterizing nematode communities at higher taxonomic levels [16]. Altogether, the variability of each genetic marker offers different resolution for phylogenetic analyses and practicability for diagnosis of nematodes [15].

Some challenges remain unsolved in the genetic analysis of nematodes. For instance, species hybrids have been initially interpreted as sequencing artifacts when discordances are detected during phylogenetic analyses using nuclear and mitochondrial genes [17]. Moreover, cryptic diversity identification in morphologically undistinguishable specimens relies on nucleotide differences among taxa and sequence availability in databases [18]. Therefore, cryptic diversity is accurately detected only when enough sequences are compared and when appropriate nuclear, ribosomal or mitochondrial markers are employed.

The aim of this study was to analyze the resolution and performance of six different ribosomal and mitochondrial loci (18S, ITS-1, ITS-2, *cox1*, 12S and 16S) for the interspecies differentiation of nematodes of the families Ascarididae, Ancylostomatidae and Onchocercidae of human and veterinary importance. Importantly, we estimated the number of available sequences on databases for each species included in the study to contextualize their usefulness in the clinical and research scenario. In this way, we evaluated the practicality of ribosomal and mitochondrial markers for the identification of worms as well as their ease of analysis and taxonomic resolution.

## Methods

Six genetic markers from representative nematode species of medical and veterinary importance of the families Ascarididae, Onchocercidae and Ancylostomatidae were analyzed in this study (Fig. 1). Accordingly, five



**Fig. 1** Ribosomal and mitochondrial sequences used for the phylogenetic analysis of 30 species of the Ascarididae, Ancylostomatidae and Onchocercidae families. The number of sequences is color coded from darkest (n=5) to lightest (n=0) according to the marker: 18S in blue, ITS1 in purple, ITS2 in red, cox1 in yellow, 12S in green and 16S in turquoise. This figure was created using Biorender.com

representative full-length or partial sequences of > 300 bp nuclear ribosomal regions ITS-1, ITS-2 and 18S rRNA and mitochondrial genes *cox1*, 12S and 16S rRNA were downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov/>) and aligned in MEGA X using the MUSCLE tool. Chosen sequences had to overlap with the consensus sequence block in at least 80% of length of the

block to be included in the analysis. To take into account intraspecies variability, up to five sequences belonging to different hosts or geographical locations were included in the analysis. The accession numbers of the sequences included in this study are available in Additional file 1: Table S1.

The best nucleotide substitution model for each loci and family was computed with jModelTest software [19] based on the Bayesian inference criterion calculations (Additional file 2: Table S2). Bayesian inference (BI) phylogenetic trees were constructed using the BEAST2 package (<https://www.beast2.org/>) with  $10^8$  MCMCs and a tree burn-in percentage of 10%. Then, the program Interactive Tree of Life (iTOL) (<https://itol.embl.de/>) was used to edit the resulting trees. Afterwards, we elaborated a color-coded matrix based on sequence pairwise identity using the program Sequence Demarcation Tool version 1.2 (SDTv1.2) [20], with alignments as input files and removing the option for calculating a neighbor-joining tree.

Violin plots of genetic p-distances were constructed using the R package ggplot2 [21]. Pairwise nucleotide p-distances were calculated in MEGA X. Data and script used for this purpose are available in the Additional file 3. The Gaussian distribution of data was verified using the Kolmogorov-Smirnov test to determine significance between pairwise p-distance values between markers. Then, a Kruskal-Wallis analysis for non-parametric data was performed followed by Dunn's multiple comparison test, and significance was considered when  $P < 0.01$ . These analyses were run in GraphPad Prism version 10.

#### Sequence availability

The total number of sequences from the abovementioned species and genetic markers were calculated using the NCBI database and BLAST tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The search was done to assess the practicality of each marker in the context of a diagnostic challenge. BLAST search was done with a complete sequence for each species. If there was no reference sequence for a particular locus, the longest sequence was selected. To retrieve all sequences available in the NCBI database, all sequences obtained from the blastn search of the reference sequences were downloaded. In addition, a manual search using key words for each genetic marker and species was also performed to retrieve more divergent sequences. Afterwards, we proofread all sequences and maintained only those with a length  $> 300$  bp and  $> 97\%$  percentage of identity with the reference sequence.

## Results

### Analysis of 18S rRNA

The analyses for the 18S rRNA region showed a minimum paired identity  $> 96.3\%$  for the three studied families (Fig. 2). Accordingly, the lowest nucleotide similarity within Ascarididae species was  $98.3\%$ , in Ancylostomatidae was  $99.60\%$  and for Onchocercidae members was  $96.3\%$ .

The BI analysis revealed low posterior probabilities (PP) in the branches separating several species of the Ascarididae and Ancylostomatidae families, indicating sequences could not be resolved and separated with enough resolution. Therefore, the phylogenetic analysis with the 18S does not present a clear differentiation of the species of these families (Fig. 2a and b). When *Toxocara canis*, *T. cati* and *T. vitulorum* were compared, nucleotide similarities were  $> 99\%$ , and all sequences retrieved from these species grouped in the same cluster with low PP in the inner ramifications. For the Ancylostomatidae family, only ten sequences could be obtained from GenBank, and these were divided in a tree branch containing the sequences of *Ancylostoma* spp. and *N. americanus*, while another tree branch contained *Uncinaria stenocephala* specimens (Fig. 2b), with low PP in inner ramifications. For the Onchocercidae, four main clusters were observed in the BI tree: one with *Diriofilaria repens*, *Loa loa* and *Onchocerca fasciata*; a second with *W. bancrofti*, *B. malayi* and a *D. repens* sequence from a dog from France; a third group with *Mansonella* spp.; a final cluster with *Diriofilaria immitis* sequences (Fig. 2c).

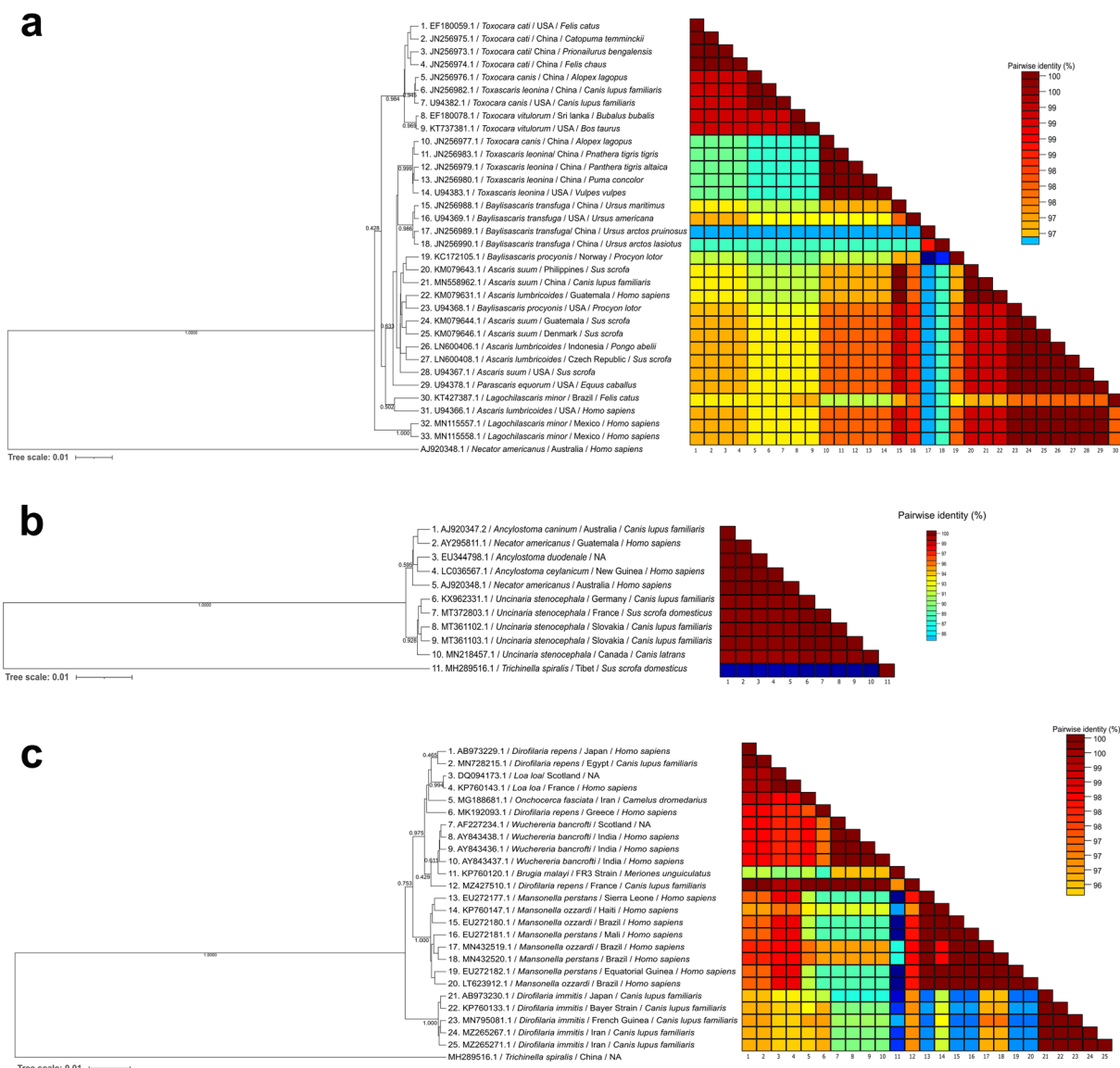
### Analysis of ITS-1 and ITS-2 ribosomal loci

The ITS regions showed high interspecies resolution with nucleotide pairwise identities  $> 51.2\%$ . The ITS-1 region had an average minimum pairwise identity of  $51.2\%$ ,  $57.9\%$  and  $58.3\%$  for the Ascarididae, Ancylostomatidae and Onchocercidae families, respectively, with complete separation of species in the BI trees (Additional file 4). Moreover, the ITS-2 showed  $52.5\%$  mean nucleotide pairwise similarity in the Ascarididae family,  $72.8\%$  for Ancylostomatidae and  $51.5\%$  in the Onchocercidae family (Fig. 3).

A total separation of species of the Ascarididae was observed in the ITS-1 and ITS-2 BI tree families, supported by PP values of 1.000 (Additional file 4 and Fig. 3). The ITS-1 and ITS-2 BI trees showed two clear clusters: one containing *Toxascaris leonina*, *Ascaris lumbricoides*, *A. suum*, *P. equorum*, *Baylisascaris procyonis* and *B. transfuga* sequences and a second containing *T. canis*, *T. cati* and *T. vitulorum* sequences. Remarkably, *A. lumbricoides* sequences from humans and *A. suum*, obtained from pigs and a human, clustered together with a minimum identity percentage of  $98.4\%$  for the ITS-1 and  $100.0\%$  for ITS-2. In addition, sequences of these two ascarids were intermixed in the cluster in the ITS-2 BI tree (Fig. 3a).

Four main groups could be identified in the ITS-1 and ITS-2 phylogenetic trees of the Ancylostomatidae family: one represented by *Ancylostoma* spp., a second represented by *Uncinaria sanguinis*, one represented by *Bunostomum* spp. and a group represented by *Necator*





**Fig. 2** Phylogenetic Bayesian inference analysis of the nuclear 18S rRNA of selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c). Nucleotide pairwise identity is denoted in the heatmap next to each taxa name

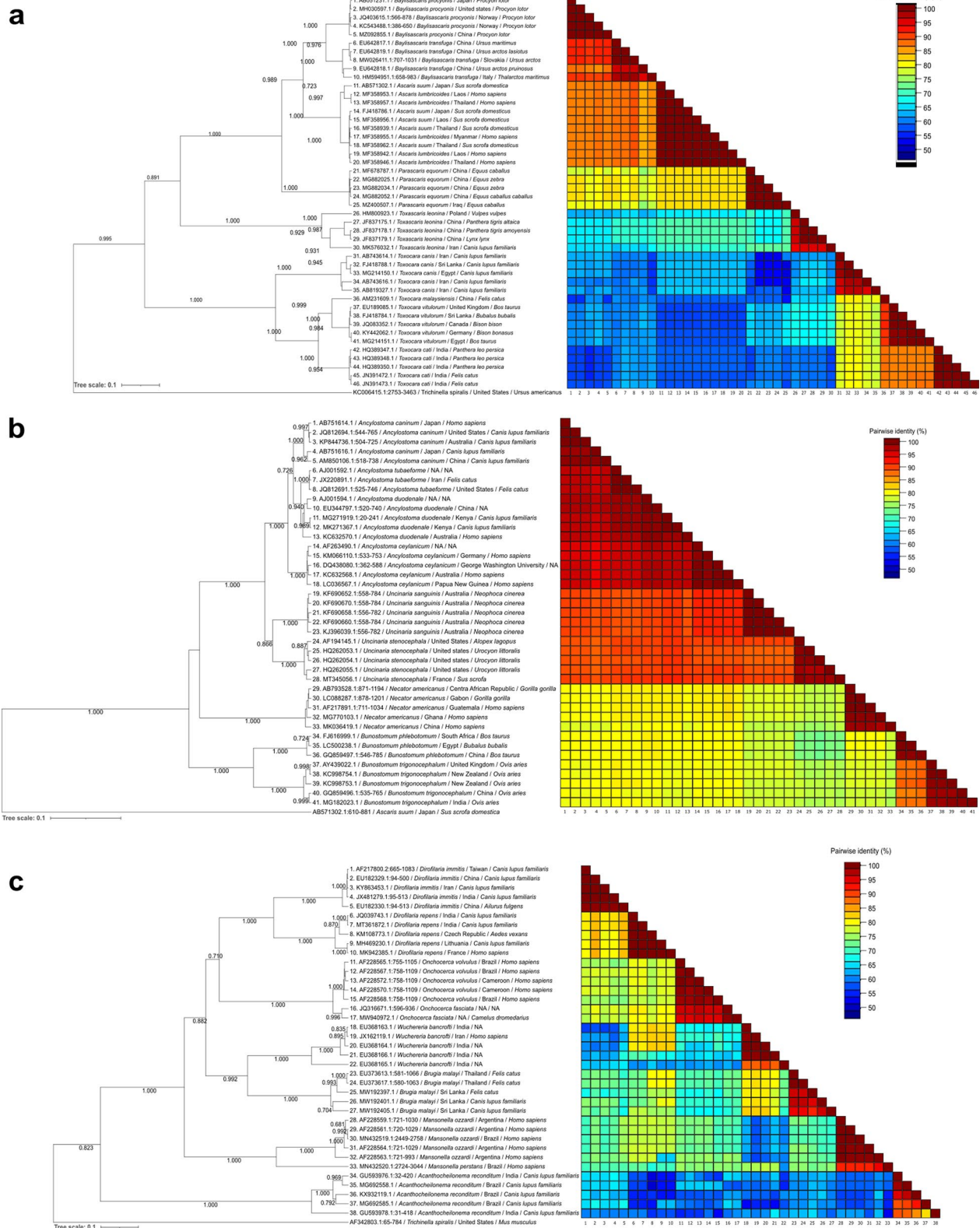
*americanus* sequences, all with PP=1.000. Moreover, each species within each cluster was separated between them, showing a clear interspecies resolution.

For the Onchocercidae family, four sequence groups were observed in the BI tree of the ITS-1 region. Accordingly, one group contained *Mansonella* spp. sequences, the second contained *Onchocerca* spp., the third contained *W. bancrofti* and *B. malayi* and, lastly, a cluster contained *Dirofilaria* spp. and *L. loa* sequences (Additional file 4). Each species was separated with high nucleotide difference, except for *W. bancrofti* and *B. malayi*, whose differences were approximately 10%. ITS-2 phylogenetic tree built with sequences of this

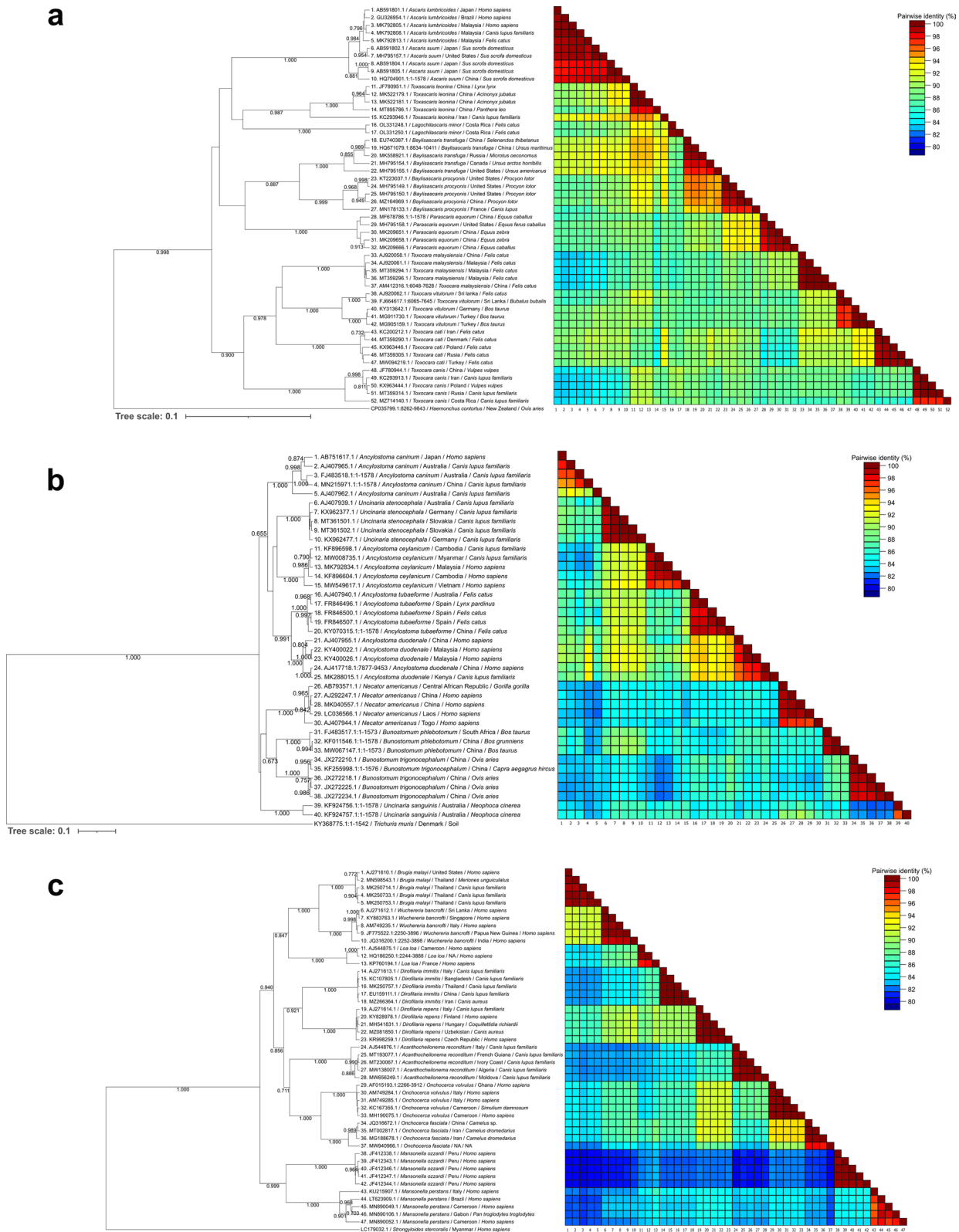
family exhibited the same clustering as ITS-1, with the addition of *Acanthocheilonema reconditum* sequences, which were located next to the *Mansonella* spp. group. Although four ITS-1 sequences of *A. reconditum* were available in GenBank, these did not meet the criteria to be included in the BI analysis.

**cox1 mitochondrial gene**

Average pairwise nucleotide similarities in the *cox1* gene fragment within the Ascarididae, Ancylostomatidae and Onchocercidae families were 85.4%, 83.0% and 80.5%, respectively (Fig. 4). All species used in the analysis were successfully resolved in the phylogenetic tree for ascarids



**Fig. 3** Phylogenetic Bayesian inference analysis of ITS2 loci of selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c). Nucleotide pairwise identity is denoted in the heatmap next to each taxa name



**Fig. 4** Phylogenetic Bayesian inference analysis of *cox1* markers of selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c). Nucleotide pairwise identity is denoted in the heatmap next to each taxa name



and showed a similar topology with the BI phylogenetic tree using the ITS-1 and ITS-2. Accordingly, three main clusters were observed: one with *A. lumbricoides*, *A. suum*, *T. leonina* and *Lagochilascaris minor*, a second group with *B. transfuga*, *B. procyonis* and *P. equorum* and a final one with *Toxocara* spp. Finally, as previously observed in the ITS-1 and ITS-2 trees, *A. lumbricoides* and *A. suum* clustered together, resulting in very high paired identity values, with the lowest value being 96.9% and the highest 100.0% (Fig. 4a). Moreover, two groups were observed in the Ancylostomatidae family tree, namely one with *Ancylostoma* spp. and *U. stenocephala* and another cluster with *N. americanus* and *Bunostomum* spp. (Fig. 4b). Phylogenetic analysis of the Onchocercidae sequences showed a marked divergence from *Mansonella* spp. regarding the other species of the family with high PP. In addition, *Acanthocheilonema reconditum* sequences were grouped with those of *Onchocerca* spp., in contrast to the topologies for the ITS-1 and ITS-2 trees, which were separated into different groups (Fig. 4c).

### 12S and 16S mitochondrial genes

Phylogenetic interspecies resolution was obtained in the analyzed 12S and 16S fragments (Additional files 5 and 6, respectively). The average pairwise nucleotide similarity in the 12S was 75.9% within the Ascarididae family, 84.7% for Ancylostomatidae and 81.5% within the Onchocercidae. On the other hand, the average pairwise nucleotide similarity for the 16S gene was 79.0%, 83.1% and 83.5% within the analyzed ascarid, ancylostomatid and onchocercid sequences. In addition, the number of sequences available for phylogenetic analysis for the 16S gene was lower than in the other loci.

Phylogenetic trees for 12S and 16S genes support the results previously obtained for *A. lumbricoides* and *A. suum*, which grouped together with a percentage of identity for the 16S sequences  $\geq 97.0\%$ . Furthermore, *A. lumbricoides*, *A. suum*, *P. equorum*, *B. procyonis*, *B. transfuga* and *T. leonina* were grouped in one of the branches of the tree, while *Toxocara* spp. sequences were clustered in a separate cluster with high PP and pairwise identities ranging from 77.0 to 79.0% compared to sequences of other ascarids.

In the 12S phylogenetic tree, two main branches could be observed, one with *N. americanus* sequences and the second with *Bunostomum* spp., *Ancylostoma* spp. and *Uncinaria* spp. The tree topology for 16S sequences varied in that *Bunostomum* spp. sequences were separated from the other analyzed species, whereas *N. americanus* sequences were closer to *Ancylostoma* spp.

The phylogenetic trees of the 12S and 16S of the Onchocercidae sequences revealed a different topology

to the one observed in the *cox1* gene fragment. For the 12S tree, three groups were observed: a branch with *Onchocerca* spp. and *Dirofilaria* spp., a second group with *B. malayi*, *W. bancrofti*, *L. loa* and *Mansonella* spp., and a third cluster with *A. reconditum*. In the 16S phylogenetic tree, fewer sequences could be analyzed according to the inclusion criteria. Nevertheless, a group with *B. malayi*, *W. bancrofti*, *L. loa* and *O. volvulus* was obtained; a separate cluster was observed with *Dirofilaria* spp. and a single *M. perstans* sequence was derived from the tree. Therefore, *D. immitis* and *D. repens* were grouped together in both 12S and 16S genes, and the same was obtained with *B. malayi* and *W. bancrofti*.

### Nucleotide differences between genes of each family

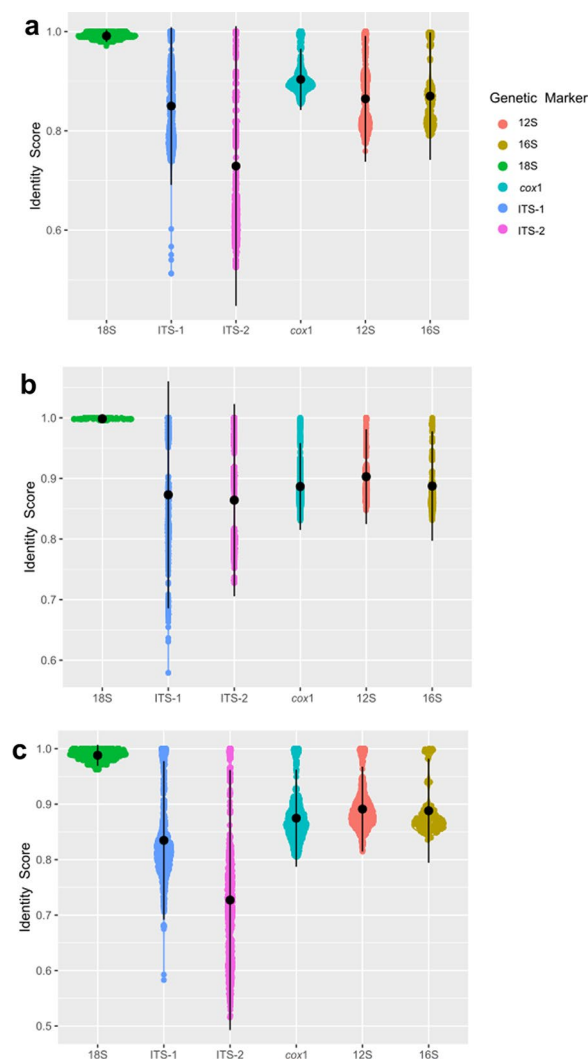
Significant pairwise nucleotide p-distances in all markers were obtained for each family (all  $P < 0.0001$ , Kruskal-Wallis statistic = 2023, 208 and 1900 for Ascarididae, Ancylostomatidae and Onchocercidae families, respectively) (Fig. 5). A total of 4998 pairwise comparisons were calculated for the Ascarididae family, 2828 for the Ancylostomatidae family and 4038 for the Onchocercidae family.

The highest pairwise nucleotide similarity in the Ascarididae family was shown in the nuclear rRNA 18S fragment with an average of  $99.1 \pm 0.6\%$ , which was significantly different from the other five markers (all  $P < 0.01$ ) (Fig. 5a). Average pairwise nucleotide p-distances of the analyzed *cox1* fragments were  $90.4 \pm 3.1\%$ , whereas the pairwise nucleotide similarities in the 16S, 12S, ITS-1 and ITS-2 markers were  $87.0 \pm 6.4\%$ ,  $86.4 \pm 6.3\%$ ,  $85.0 \pm 7.9\%$  and  $72.9 \pm 14.1\%$ , respectively. Significant differences were detected when comparing the values of each marker between them, except between average pairwise p-distances of 12S and 16S, 16S and ITS-1, and 12S and ITS-1.

For the Ancylostomatidae family, the highest pairwise similarity was obtained in the 18S fragment, with an average of  $99.8 \pm 0.1\%$  similarity between sequences, followed by the 12S with  $90.3 \pm 3.9\%$ , the 16S with  $88.7 \pm 4.5\%$ , *cox1* with  $88.7 \pm 3.6\%$ , ITS-1 with  $87.3 \pm 9.4\%$  and ITS-2 with  $86.4 \pm 7.9\%$  average pairwise similarities (Fig. 5b). The average pairwise p-distance was significantly different when comparing the data of the 18S to the other five markers with all  $P < 0.01$ . Other pairwise statistical comparisons were significant except between data of 16S and *cox1*, *cox1* and ITS-1, and ITS-1 and ITS-2, whose average pairwise distance were similar between them.

Similar results to those obtained for the Ascarididae and Ancylostomatidae families were evident for the Onchocercidae family (Fig. 5c), with the nuclear rRNA 18S showing the highest pairwise nucleotide similarities ( $98.8 \pm 0.9\%$ ) with significant difference (all  $P < 0.01$ )



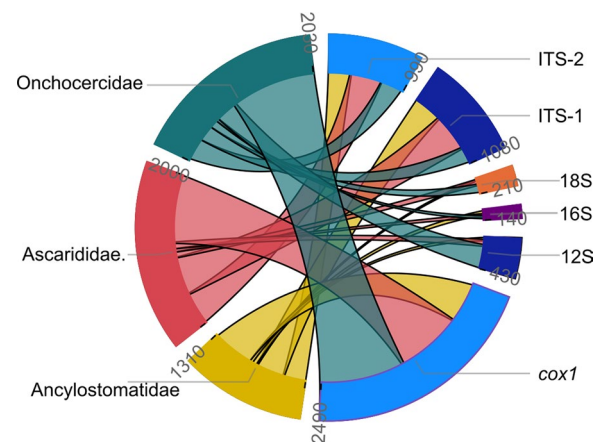


**Fig. 5** Nucleotide similarities between selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c) in each marker analyzed in this study

compared to the 12S ( $89.1 \pm 3.8\%$ ), 16S ( $88.8 \pm 4.7\%$ ), *cox1* ( $87.5 \pm 4.4\%$ ), ITS-1 ( $83.5 \pm 7.1\%$ ) and ITS-2 ( $72.7 \pm 11.7\%$ ). Significant differences were obtained in all pairwise comparisons of loci except between the 16S and 12S and the 16S and *cox1*.

#### Availability of sequences in the GenBank database

The marker with the highest representation in GenBank was the *cox1* with a total of 2491 sequences for the evaluated species, followed by the ITS-1 ( $n=1082$ ), ITS-2 ( $n=994$ ), 12S ( $n=428$ ), 18S ( $n=212$ ) and 16S ( $n=143$ ) (Fig. 6). Complete results with the number of sequences per analyzed species is available in Additional file 7. Accordingly, the analyzed species of the Ascarididae



**Fig. 6** Availability of *cox1*, 12S, 16S, ITS-1, ITS-2 and 18S sequences of selected species of veterinary and clinical importance of the families Ascarididae, Ancylostomatidae and Onchocercidae in NCBI databases. Circular Sankey chart of the number of sequences available for each marker depending on the family. Sum of sequences available for each marker by family

family had 870 *cox1* sequences deposited in GenBank, whereas 525 sequences were available for the species of the Ancylostomatidae family and 1096 for the Onchocercidae family. For the ITS-1, 498, 375 and 209 sequences were retrieved from GenBank in the studied species of the Ascarididae, Ancylostomatidae and Onchocercidae families, respectively. The marker with the least sequence availability was the 16S mitochondrial rRNA gene with 46 sequences in the analyzed species of the Ascarididae family, 56 for the Ancylostomatidae and 41 for the Onchocercidae. The numbers of sequences per species are presented in Table 1.

#### Discussion

We analyzed the interspecies resolution and sequence availability of different mitochondrial and ribosomal markers commonly used for diagnosing helminthiases of medico-veterinary significance. The present work determined that the mitochondrial gene *cox1* had the highest number of available sequences in GenBank, indicating this is a suitable marker to use when confirming the identity of an unknown specimen since many sequences are available for comparison. Moreover, this analysis included sequences from different hosts and geographical regions, thus giving a wider perspective of the genetic variability of each species. The nuclear rRNA 18S gene showed the lowest species resolution, whereas nuclear ITS-1 and ITS-2 regions showed the highest nucleotide differences between species, followed by mitochondrial gene *cox1* and mitochondrial rRNA genes 12S and 16S. Furthermore, the *cox1* gene was the most represented

**Table 1** Number of sequences retrieved from GenBank of nematodes of veterinary and clinical importance of the Ascarididae, Ancylostomatidae and Onchocercidae families

Family	Species	18S	ITS1	ITS2	cox1	12S	16S
Ascarididae	<i>Ascaris lumbricoides</i>	7	129	27	238	18	17
	<i>Ascaris suum</i>	17	44	19	114	12	7
	<i>Toxascaris leonina</i>	9	66	62	76	13	5
	<i>Baylisascaris transfuga</i>	4	15	19	25	10	2
	<i>Baylisascaris procyonis</i>	4	47	13	165	4	2
	<i>Parascaris equorum</i>	1	49	53	29	4	3
	<i>Toxocara vitulorum</i>	4	17	15	7	6	1
	<i>Toxocara malaysiensis</i>	0	0	1	9	2	2
	<i>Toxocara cati</i>	17	82	135	93	6	2
	<i>Toxocara canis</i>	12	49	75	103	11	5
	<i>Lagochilascaris minor</i>	5	0	0	11	2	0
Ancylostomatidae	<i>Ancylostoma caninum</i>	2	181	96	52	4	4
	<i>Ancylostoma tubaeforme</i>	0	22	3	20	2	2
	<i>Ancylostoma duodenale</i>	1	15	5	23	3	6
	<i>Ancylostoma ceylanicum</i>	1	67	58	128	4	3
	<i>Uncinaria sanguinis</i>	0	22	22	59	3	3
	<i>Uncinaria stenocephala</i>	16	0	24	21	15	5
	<i>Bunostomum phlebotomum</i>	0	1	3	4	4	4
	<i>Bunostomum trigenocephalum</i>	0	5	8	41	3	3
Onchocercidae	<i>Necator americanus</i>	4	62	72	177	5	26
	<i>Brugia malayi</i>	3	62	36	17	7	3
	<i>Wuchereria bancrofti</i>	5	7	5	499	7	5
	<i>Loa loa</i>	2	19	0	3	5	1
	<i>Dirofilaria immitis</i>	61	34	118	194	124	9
	<i>Dirofilaria repens</i>	25	4	31	319	114	7
	<i>Acanthocheilonema reconditum</i>	0	1	45	25	15	0
	<i>Onchocerca volvulus</i>	0	13	18	8	14	13
	<i>Onchocerca fasciata</i>	1	2	2	4	5	2
	<i>Mansonella ozzardi</i>	4	13	28	10	4	0
<i>Mansonella perstans</i>	7	54	1	17	2	1	

marker in GenBank, with almost the same number of sequences of 18S, ITS-1, ITS-2, 12S and 16S summed together for the analyzed species.

The results obtained for the 18S gene of the Ascarididae, Ancylostomatidae and Onchocercidae families indicate that this marker does not separate between close species, since nucleotide similarity values between 98.8 and 99.9% were obtained. The analyzed species were intermixed in the phylogenetic trees of the three families, especially between *A. lumbricoides* and *A. suum*, *Toxocara* spp. with *T. leonina*, *W. bancrofti* and *B. malayi*, and among *Mansonella* spp. Therefore, this marker may resolve species at higher taxonomic levels [15, 22, 23], which may be useful for describing nematode communities in soil [24, 25] or biological animal samples [26] using 18S-based metagenomics. However, it does not provide the necessary resolution for identifying unknown

specimens in a clinical setting [27]. For instance, an unknown specimen may be erroneously classified if only an 18S fragment is analyzed, such as the case of *L. minor* from cats, which showed 99% homology to *A. suum* [27]. For these cases, the analysis of several and longer DNA fragments is recommended, together with the close morphological examination of specimens. The low interspecies difference in the 18S may be explained by the concerted evolution of this coding region, which homogenizes all copies in a specimen's genome [28] and secures the structure of rRNA functional domains, as opposed to the non-coding ITS regions [29].

The ITS-1 and ITS-2 regions yielded the highest percentage of sequence variability and separated all species in phylogenetic analyses except *A. suum* and *A. lumbricoides*, which were intermixed and *Ancylostoma* spp. that showed lower nucleotide differences between them.

Although the ITS regions are part of the nuclear rRNA operon, these loci have a higher nucleotide substitution rate than the 18S [29], which eases differentiation at the species level [15]. This characteristic was confirmed in the present study and has been observed for nematodes of other families like Spirocercidae [30], Strongylidae [31] and Trichuridae [32], to name some examples. Nucleotide differences in these loci are so high that they may even exhibit a high genetic intra-individual variation, i.e. a single specimen with different copies of ITS, such as the case of the dog carcinogenic nematode *Spirocerca lupi* [30], the whipworms *Trichuris* spp. [33] or the causing agent of cystic echinococcosis, *Echinococcus granulosus* [34]. In this study, ITS-1 and ITS-2 proved to be useful markers in helminth identification applications. Nevertheless, the high nucleotide difference between species complicated subsequent analyses, such as sequence alignment and phylogenetic reconstructions when using large datasets [15]. For this reason, ideal diagnostic markers would be those showing enough interspecies nucleotide differences but easy to handle for beginners.

The three markers of mitochondrial origin analyzed in this work had a similar performance in terms of their sequence resolution in phylogenetic analyses and the pairwise nucleotide differences. It has been argued that mitochondrial rRNA 12S and 16S are less variable than *cox1*. However, we did not observe significant statistical differences in the percentage of nucleotide variation in the three markers [15]. The *cox1* gene has been used in DNA barcoding and is currently considered the universal genetic marker for this application [35]. DNA barcoding uses a small consensus DNA fragment of the *cox1* to globally identify a large number of organisms from different geographical locations, hosts or genetic subclassifications [35–37]. Mitochondrial markers have several advantages compared to those of nuclear origin. For example, mitochondrial markers do not undergo recombination and have a higher mutation rate than nuclear rRNA 18S, 5.8S or 28S [38]. Therefore, mitochondrial genes are useful in resolving organisms at lower taxonomic levels to the species level and thus may be useful for metabarcoding purposes.

Although mitochondrial markers have great variability, it has been argued that the design of universal primers targeting mitochondrial regions may be challenging [39]. However, previous studies have successfully amplified consensus *cox1* sequences from different nematode families [40] and have described free-living marine nematode communities [41]. Moreover, a study of seven nematode genera across different families showed an interspecies mean p-distance or nucleotide difference in the *cox1* between 10 and 25%, whereas intraspecific pairwise

nucleotide differences ranged from 0.5 to 3.5% [39]. Those results resemble the ones obtained in the present study. Altogether, our results support the applicability of *cox1* for metabarcoding studies.

The number of *cox1* sequences available in GenBank exceeded those of the other five markers according to the employed mining pipeline. The present study analyzed this aspect for the first time, thus giving a different perspective regarding the usefulness of each locus in a diagnostic context. The larger number of *cox1* sequences may be explained by the high sensitivity of the primers used for amplifying this gene, the widespread use of certain primers targeting the *cox1* and the efforts of research groups worldwide in using and depositing barcodes of their specimens, thus including all possible biodiversity [39]. Interspecies resolution obtained in the 12S and 16S rRNA mitochondrial genes were comparable to those of the *cox1* and therefore may have similar utility in metabarcoding analyses. However, the use of 12S and 16S and their applicability for phylogenetic, taxonomic and barcoding analysis would be limited by the low number of sequences in databases. A similar situation was observed with ITS regions, which provide larger nucleotide differences between species compared to nucleotide markers as obtained before [15], but their lower availability would limit their use for diagnostic or metabarcoding purposes. Therefore, sequencing and deposition of *cox1*, 12S, 16S, ITS-1 and ITS-2 sequences from nematode specimens are strongly encouraged to include the most comprehensive genetic information, which will be useful in future studies.

Interestingly, *A. lumbricoides* and *A. suum* sequences were intermixed in all analyzed markers, thus raising the long-standing question of whether these two taxonomic units actually represent the same species [42]. Cryptic diversity is defined as two morphologically indistinguishable but genetically different entities [43]. Additionally, other parameters have been used to infer a cryptic species including host preference, geographic distribution, resource use and reproductive behavior [18]. Accordingly, both ascarids have been proposed as cryptic species since *A. suum* was named as a pig roundworm by Goeze (1782), even though it did not present any morphological differences from *A. lumbricoides*. The present study showed that *A. suum* and *A. lumbricoides* clustered together in the phylogenetic trees of the different genes and had a pairwise nucleotide identity between 97 and 100%. In fact, genomic analyses of *A. lumbricoides* obtained from humans of a village without pig husbandry showed a high heterozygosity and hybridization of *A. lumbricoides* and *A. suum*, suggesting a potential interbred *Ascaris* complex infecting humans [44]. In any case,

an allele-specific PCR using the ITS-1 of *A. suum* and *A. lumbricoides* showed good resolution for discriminating both ascarids, which may become useful for epidemiological and control purposes [45].

The selection of markers used in the present study was not exhaustive, and they were chosen according to their current use for diagnostic purposes. Other mitochondrial genes such as the *nad1* and *cytb* have been evaluated before [15], rendering similar interspecies variation as obtained herein for *cox1*, 12S and 16S analyses. Additionally, representative nuclear and mitochondrial sequences of nematodes, trematodes and cestodes of clinic and veterinary importance have been evaluated before. Nevertheless, we focused on 30 species of three nematode families of medico-veterinary importance, rendering up to five sequences from each species when available. In this way, genetic variation arising from geographic or host origin could be included herein, which has not been studied before to our knowledge. Importantly, these kind of studies are necessary to guide basic and medical researchers in their decision-making when unknown worms arrive at their laboratories or clinics.

Even though molecular diagnosis of nematodes may be objective and straightforward [46], a proper nematode identification should include morphological analysis of specimens [14]. Molecular data complement morphological observations of worms, which may lead to the identification of cryptic diversity *sensu lato* or *sensu stricto*, and the reclassification of certain taxa or the description of new species [18]. This integrative taxonomic approach has been suggested before for filarioids [14] and has proven essential for the description of the true biodiversity of a genus, family or phyla. Furthermore, incorrect deposition of genetic data has been demonstrated in other invertebrate orders [47] and for the dyphyllobothriean cestodes of the genus *Spirometra* [48]. This highlights the responsibility of researchers in the selection of methods and the interpretation of their results.

## Conclusions

Phylogenetic relationships and minimal pairwise nucleotide identities were analyzed in six different molecular markers from 30 species of the Ascarididae, Ancylostomatidae and Onchocercidae families. Importantly, we assessed, for the first time to our knowledge, the availability of those sequences in databases as a complement for choosing a molecular marker in a clinical or research context. Results showed that *cox1* is a useful marker for diagnostic or metabarcoding purposes due to the high interspecies resolution demonstrated in phylogenetic analyses and the higher number of sequences available in databases. This will enable the accurate identification of an unknown specimen when comparing their associated

sequences to those deposited in GenBank. Finally, we recommend the combined analysis of ITS regions, other mitochondrial genes and morphological analysis of specimens whenever possible.

## Abbreviations

*cox1* Cytochrome oxidase subunit 1  
ITS Internal transcribed spacer

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13071-023-06113-4>.

**Additional file 1: Table S1.** Accession numbers of sequences used in the study.

**Additional file 2: Table S2.** Best nucleotide substitution models for each marker and family used in the analysis.

**Additional file 3.** Script used for constructing violin plots with nucleotide p-distances.

**Additional file 4.** Phylogenetic Bayesian inference analysis of ITS1 loci of selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c). Nucleotide pairwise identity is denoted in the heatmap next to each taxa name.

**Additional file 5.** Phylogenetic Bayesian inference analysis of 12S genes of selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c). Nucleotide pairwise identity is denoted in the heatmap next to each taxa name.

**Additional file 6.** Phylogenetic Bayesian inference analysis of 16S loci of selected species of veterinary and clinic importance of the families Ascarididae (a), Ancylostomatidae (b) and Onchocercidae (c). Nucleotide pairwise identity is denoted in the heatmap next to each taxa name.

**Additional file 7.** PowerBI dashboard of 18S rRNA, ITS-1, ITS-2, *cox1*, 12S and 16S sequences available in GenBank for 30 species of the Ascarididae, Ancylostomatidae and Onchocercidae families. <https://app.powerbi.com/view?r=eyJrjoiZGNhN2MwY2Etyj1OS00NjFjLWlzZWItMTk0NGlzMzA1NTM2liwidCl6ImU3OTg0Y2FjLT11NDMtNGY4OC04Zjk3LTk1MjQzMzVlNmJjNCslmMiOjR9>.

## Acknowledgement

The authors would like to thank lab members for their critical comments to this research work.

## Author contributions

MJM-A, CP-S and AR conceived the study, MJM-A, CP-S, PA-S, JQ, JQ-R, RG and EJR analyzed the data and prepared the figures, MJM-A, CP-S and AR wrote the manuscript draft. All authors read and approved the final version of the manuscript.

## Funding

This work was supported by the Vicerrectoría de Investigación of the University of Costa Rica by Grant C2064 granted to AR.

## Availability of data and materials

Analysis of sequence availability is publicly available at <https://app.powerbi.com/view?r=eyJrjoiZGNhN2MwY2Etyj1OS00NjFjLWlzZWItMTk0NGlzMzA1NTM2liwidCl6ImU3OTg0Y2FjLT11NDMtNGY4OC04Zjk3LTk1MjQzMzVlNmJjNCslmMiOjR9>.

## Declarations

### Ethics approval and consent to participate

Not applicable.



**Consent for publication**

Not applicable.

**Competing interests**

Authors declare they have no competing interests.

**Author details**

<sup>1</sup>Laboratorio de Helmintología, Departamento de Parasitología, Facultad de Microbiología, University of Costa Rica, San José, Costa Rica. <sup>2</sup>Centro de Investigación en Enfermedades Tropicales, University of Costa Rica, San José, Costa Rica. <sup>3</sup>National Reference Center for Bacteriology, Costa Rican Institute for Research and Teaching in Nutrition and Health (INCIENSA), Tres Rios, Costa Rica. <sup>4</sup>Ross University School of Veterinary Medicine, West Farm, Basseterre, Saint Kitts and Nevis.

Received: 24 October 2023 Accepted: 26 December 2023

Published online: 20 February 2024

**References**

- Chandy A, Thakur AS, Singh MP, Manigauha A. A review of neglected tropical diseases: filariasis. *Asian Pac J Trop Med*. 2011;4:581–6.
- Wang J. Genomics of the parasitic nematode *ascaris* and its relatives. *Genes*. 2021;12.
- Aschale Y, Worku L, Addisu A, Alemu M, Alemu A. Comparison of direct wet mount, Kato-Katz and formol ether sedimentation technique for the diagnosis of hookworm infection in Debre Elias Woreda, Northwest Ethiopia. *Trop Doctor*. 2021;51:170–4.
- Zibaei M, Nosrati MRC, Shadnoosh F, Houshmand E, Karami MF, Rafsanjani MK, et al. Insights into hookworm prevalence in Asia: a systematic review and meta-analysis. *Trans R Soc Trop Med Hyg*. 2020;114:141–54.
- Figueredo Duarte Moraes M, de Souza Pollo A, Lux Hoppe EG. Filarids (Spirurida: Onchocercidae) in wild carnivores and domestic dogs from the Brazilian Atlantic forest. *PLoS Negl Trop Dis*. 2022;16:e0010213.
- Blaxter M, Koutsovoulos G. The evolution of parasitism in Nematoda. *Parasitology*. 2015;142:S26–39.
- Azzopardi KI, Hardy M, Baker C, Bonnici R, Llewellyn S, McCarthy JS, et al. Detection of six soil-transmitted helminths in human stool by qPCR- a systematic workflow. *PLoS ONE*. 2021;16:e0258039.
- French SK, Pearl DL, Peregrine AS, Jardine CM. Baylisascaris procyonis infection in raccoons: a review of demographic and environmental factors influencing parasite carriage. *Vet Parasitol Reg Stud Rep*. 2019;16:100275.
- Rostami A, Riahi SM, Hofmann A, Ma G, Wang T, Behniafar H, et al. Global prevalence of *Toxocara* infection in dogs. *Adv Parasitol*. 2020;109:561–83.
- Mota KCP, Grama DF, Fava NMN, Ungari LP, Faria ESM, Cury MC. Distribution and risk factors of Ascarididae and other geohelminths in the soil of Uberlândia, Minas Gerais, Brazil. *Rev Inst Med Trop São Paulo*. 2018;60:e17.
- Nikolay B, Brooker SJ, Pullan RL. Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. *Int J Parasitol*. 2014;44:765–74.
- Jourdan PM, Lambertson PHL, Fenwick A, Addiss DG. Soil-transmitted helminth infections. *Lancet*. 2018;391:252–65.
- Magnis J, Lorentz S, Guardone L, Grimm F, Magi M, Naucke TJ, et al. Morphometric analyses of canine blood microfilariae isolated by the Knott's test enables *Dirofilaria immitis* and *D. repens* species-specific and *Acanthocheilonema* (syn. *Dipetalonema*) genus-specific diagnosis. *Parasit Vectors*. 2013;6:48.
- Ferri E, Barbuto M, Bain O, Galimberti A, Uni S, Guerrero R, et al. Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). *Front Zool*. 2009;6.
- Chan AHE, Chaisiri K, Saralamba S, Morand S, Thaenkham U. Assessing the suitability of mitochondrial and nuclear DNA genetic markers for molecular systematics and species identification of helminths. *Parasit Vectors*. 2021;14:233.
- Hwang UW, Kim W. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *Korean J Parasitol*. 1999;37:215–28.
- Chan AHE, Chaisiri K, Morand S, Saralamba N, Thaenkham U. Evaluation and utility of mitochondrial ribosomal genes for molecular systematics of parasitic nematodes. *Parasit Vectors*. 2020;13:364.
- Chaves-Gonzalez LE, Morales-Calvo F, Mora J, Solano-Barquero A, Verocai GG, Rojas A. What lies behind the curtain: cryptic diversity in helminth parasites of human and veterinary importance. *Curr Res Parasitol Vector Borne Dis*. 2022;2:100094.
- Posada D. Selection of models of DNA evolution with jModelTest. *Methods Mol Biol*. 2009;537:93–112.
- Muhire BM, Varsani A, Martin DP. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE*. 2014;9:e108277.
- Wickham H. ggplot2: elegant graphics for data analysis. 2 edn: Springer Cham; 2016.
- Brindley PJ, Mitreva M, Ghedin E, Lustigman S. Helminth genomics: the implications for human health. *PLoS Negl Trop Dis*. 2009;3:e538.
- Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, et al. A molecular evolutionary framework for the phylum Nematoda. *Nature*. 1998;392:71–5.
- Ahmed M, Back MA, Prior T, Karszen G, Lawson R, Adams I, et al. Meta-barcoding of soil nematodes: the importance of taxonomic coverage and availability of reference sequences in choosing suitable marker(s). *Metabarcoding & Metagenomics*. 2019;3:77–99.
- Sikder MM, Vestegard M, Sapkota R, Kyndt T, Nicolaisen M. A novel metabarcoding strategy for studying nematode communities. *BioRxiv*. 2020.
- Tanaka R, Hino A, Tsai IJ, Palomares-Rius JE, Yoshida A, Ogura Y, et al. Assessment of helminth biodiversity in wild rats using 18S rDNA based metagenomics. *PLoS ONE*. 2014;9:e110769.
- Solano-Barquero A, Estrada A, Medaglia A, Montenegro VM, Rojas A. Emerging *Lagochilascaris minor* infections in domestic cats from Costa Rica: a zoonotic threat for the region. *Vet Parasitol Reg Stud Reports*. 2022;36:100797.
- Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*. 2007;175:477–85.
- Qing X, Bik H, Yergaliyev TM, Gu J, Fonderie P, Brown-Miyara S, et al. Widespread prevalence but contrasting patterns of intragenomic rRNA polymorphisms in nematodes: implications for phylogeny, species delimitation and life history inference. *Mol Ecol Resour*. 2020;20:318–32.
- Rojas A, Dvir E, Farkas R, Sarma K, Borthakur S, Jabbar A, et al. Phylogenetic analysis of *Spirocerca lupi* and *Spirocerca vulpis* reveal high genetic diversity and intra-individual variation. *Parasit Vectors*. 2018;11:639.
- Sultana Y, Kong F, Mukutmoni M, Fahria L, Begum A, Lee R. Internal transcribed spacer region 1 as a promising target for detection of intra-specific polymorphisms for *Strongyloides stercoralis*. *Trop Parasitol*. 2022;12:48–53.
- Phosuk I, Sanpool O, Thanchomnang T, Sadaow L, Rodpai R, Anamnart W, et al. Molecular identification of *Trichuris suis* and *Trichuris trichiura* Eggs in Human Populations from Thailand, Lao PDR, and Myanmar. *Am J Trop Med Hyg*. 2018;98:39–44.
- Callejon R, Halajian A, de Rojas M, Marrugal A, Guevara D, Cutillas C. 16S partial gene mitochondrial DNA and internal transcribed spacers ribosomal DNA as differential markers of *Trichuris discolor* populations. *Vet Parasitol*. 2012;186:350–63.
- Bowles J, Blair D, McManus DP. A molecular phylogeny of the genus *Echinococcus*. *Parasitology*. 1995;110:317–28.
- Hajibabaei M, Singer GA, Hebert PD, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet*. 2007;23:167–72.
- Ratnasingham S, Hebert PD. bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes*. 2007;7:355–64.
- Ratnasingham S, Hebert PD. A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE*. 2013;8:e66213.
- Yang L, Tan Z, Wang D, Xue L, Guan MX, Huang T, et al. Species identification through mitochondrial rRNA genetic analysis. *Sci Rep*. 2014;4:4089.
- Goncalves LT, Bianchi FM, Deprá M, Calegaro-Marques C. Barcoding a can of worms: testing cox1 performance as a DNA barcode of Nematoda. *Genome*. 2021;64:705–17.
- Prosser SW, Velarde-Aguilar MG, Leon-Regagnon V, Hebert PD. Advancing nematode barcoding: a primer cocktail for the cytochrome c oxidase

- subunit I gene from vertebrate parasitic nematodes. *Mol Ecol Resour.* 2013;13:1108–15.
41. Derycke S, Vanaverbeke J, Rigaux A, Backeljau T, Moens T. Exploring the use of cytochrome oxidase c subunit 1 (COI) for DNA barcoding of free-living marine nematodes. *PLoS ONE.* 2010;5:e13716.
  42. da Silva Alves EB, Conceicao MJ, Leles D. *Ascaris lumbricoides*, *Ascaris suum*, or “*Ascaris lumbricoides*”? *J Infect Dis.* 2016;213:1355.
  43. Nadler SA, Pérez-Ponce de León G. Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology. *Parasitology.* 2011;138:1688–709.
  44. Easton A, Gao S, Lawton SP, Bennuru S, Khan A, Dahlstrom E, et al. Molecular evidence of hybridization between pig and human *Ascaris* indicates an interbred species complex infecting humans. *Elife.* 2020;9.
  45. Dos Santos TR, Furtado LFV, de Carvalho Araujo A, da Silva Medeiros C, Germano PHV, de Oliveira V, et al. Development of allele-specific PCR methodology (AS-PCR) to screening *A. lumbricoides* and *A. suum*. *Parasitol Res.* 2022;121:2389–97.
  46. Christoforou M, Orford M, Tsaltas D. Molecular diagnostic tools for nematodes. In: Manjur M, Mahamood M, editors. *Nematology—concepts, diagnosis and control.* Intech Open; 2017. p. 192.
  47. Mioduchowska M, Czyz MJ, Goldyn B, Kur J, Sell J. Instances of erroneous DNA barcoding of metazoan invertebrates: are universal cox1 gene primers too “universal”? *PLoS ONE.* 2018;13:e0199609.
  48. Kuchta R, Kolodziej-Sobocinska M, Brabec J, Mlocicki D, Salamatin R, Scholz T. Sparganosis (*Spirometra*) in Europe in the Molecular Era. *Clin Infect Dis.* 2021;72:882–90.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.