NRC Research Press

ARTICLE

Barcoding a can of worms: testing *cox1* performance as a DNA barcode of Nematoda

Leonardo Tresoldi Gonçalves, Filipe Michels Bianchi, Maríndia Deprá, and Cláudia Calegaro-Marques

Abstract: Accurate taxonomic identifications and species delimitations are a fundamental problem in biology. The complex taxonomy of Nematoda is primarily based on morphology, which is often dubious. DNA barcoding emerged as a handy tool to identify specimens and assess diversity, but its applications in Nematoda are incipient. We evaluated cytochrome *c* oxidase subunit I (*cox1*) efficiency as a DNA barcode for nematodes scrutinising 5241 sequences retrieved from BOLD and GenBank. The samples included genera with medical, agricultural, or ecological relevance: *Anguillicola*, *Caenorhabditis*, *Heterodera*, *Meloidogyne*, *Onchocerca*, *Strongyloides*, and *Trichinella*. We assessed *cox1* performance through barcode gap and Probability of Correct Identification (PCI) analyses, and estimated species richness through Automatic Barcode Gap Discovery (ABGD). Each genus presented distinct gap ranges, mirroring the evolutionary diversity within Nematoda. Thus, to survey the diversity of the phylum, a careful definition of thresholds for lower taxonomic levels should be considered. PCIs were around 70% for both databases, highlighting operational biases and challenges in nematode taxonomy. ABGD inferred higher richness than the taxonomic labels informed by databases. The prevalence of specimen misidentifications and dubious species delimitations emphasise the value of integrative approaches to nematode taxonomy and systematics. Overall, *cox1* is a relevant tool for integrative taxonomy of nematodes.

Key words: BOLD, COI, GenBank, integrative taxonomy, specimen identification.

Résumé : La justesse des identifications taxonomiques et de la délimitation des espèces constitue un problème fondamental en biologie. La taxonomie complexe des Nematoda est fondée principalement sur l'étude de la morphologie, laquelle est souvent douteuse. Le codage à barres de l'ADN a émergé comme moyen pratique pour identifier des spécimens et pour mesurer la diversité, mais son emploi au sein des Nematoda est encore embryonnaire. Les auteurs ont évalué la performance de la sous-unité I de la cytochrome *c* oxydase (*cox1*) à titre de code-barre de l'ADN chez les nématodes en examinant 5241 séquences obtenues des bases de données BOLD et GenBank. Les échantillons incluaient des genres d'intérêt médical, agricole ou écologique : *Anguillicola*, *Caenorhabditis*, *Heterodera*, *Meloidogyne*, *Onchocera*, *Strongyloides* et *Trichinella*. Les auteurs ont mesuré la performance de *cox1* via des analyses de l'écart entre les codes-barre et de la probabilité d'identification correcte (PCI pour "Probability of Correct Identification"), en plus d'estimer la richesse en espèces via l'outil ABGD ("Automated Barcode Gap Discovery"). Chaque genre présentait des étendues distinctes en matière d'écart, ce qui reflète la diversité évolutive au sein des Nematoda. Il en découle que, pour mesurer la diversité au sein du phylum, il est nécessaire de procéder à une définition méticuleuse des seuils pour les niveaux taxonomiques inférieurs. Les PCI avoisinaient 70 % au sein des deux bases de données, ce qui suggère des biais opérationnels et des défis dans la taxonomie des nématodes. L'analyse ABGD a indiqué une plus grande richesse d'espèces que les identifications taxonomiques indiquées dans les bases de données. La fréquence de mauvaise identification de spécimens et de délimitations douteuses des espèces fait ressortir l'intérêt d'avoir recours à des approches intégratives pour les études de

Received 17 August 2020. Accepted 28 December 2020.

L.T. Gonçalves and C. Calegaro-Marques. Laboratório de Helmintologia, Departamento de Zoologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

F.M. Bianchi. Laboratório de Entomologia Sistemática, Departamento de Zoologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

M. Deprá. Laboratório de Drosophila, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Corresponding author: Leonardo Tresoldi Gonçalves (email: tresoldigoncalves@gmail.com).

taxonomie et de systématique chez les nématodes. Globalement, *cox1* s'est avéré un outil pertinent pour la taxonomie intégrative chez les nématodes.

Mots-clés : BOLD, COI, GenBank, taxonomie intégrative, identification de spécimens.

Introduction

The correct identification and delimitation of taxa are crucial to applied surveys, in addition to ecological, taxonomic, systematic, and evolutionary studies (Hey 2009). Exponential advances in molecular biology have mitigated a myriad of biological problems related to organism identification. The use of DNA sequences to specimen identification started in the 1980s (e.g., Kloos and Wolfshohl 1982; Rollinson et al. 1986; Gale and Crampton 1987), boosting the interest of scientists to improve molecular practices, theories, and analytical methods. DNA barcoding was formalised at the beginning of this millennium, promising precise identification of specimens through a DNA sequence fragment from a standardised region of the genome (Hebert et al. 2003). For animals, the primary DNA barcode is a 658 base pairs (bp) region of the mitochondrial gene cytochrome *c* oxidase subunit I (*cox1*) (Ratnasingham and Hebert 2007).

A recent review of interpretations and trends in DNA barcoding shows a constant rise in studies using this tool to solve different biological problems (e.g., species delimitation, species discovery, specimen identification) within distinct disciplines (DeSalle and Goldstein 2019). For a proper use of DNA barcoding, each specific evolutionary lineage demands careful preliminary analyses. In the lack of previous information about a specific taxon, researchers set threshold values a priori. Many studies use an arbitrary value between 2% and 3% for specific divergence, depending on the taxonomic group (Hebert et al. 2003; Abdo and Golding 2007; Clare et al. 2007). On the other hand, the Barcode of Life Data System (BOLD) initially assigns a 1% distance threshold, leading to the recognition of a higher number of operational taxonomic units (Ratnasingham and Hebert 2007).

Unlike fundamental thoughts concerning DNA barcoding, there is no prior reason to assume a universal fixed threshold value to sort out conspecific from heterospecific taxa. Since coalescent depths among species vary intrinsically for each lineage (Fujita et al. 2012), a fixed threshold for all organisms would generate false-positive and false-negative errors, depending on the pooled species (Goldstein et al. 2000). A shortcoming for distance-based methods is the lack of objective criteria to delineate lineages (DeSalle et al. 2005), and only the accumulation of data, their compilation in digital libraries, and their analytical interpretation can improve the detection and optimisation of an empirical threshold value for a specific taxon—preferably lineages closer to species level (e.g., genus level).

In addition to specimen identification, DNA barcoding may assist in species discovery, through the search for a "barcode gap" (Meyer and Paulay 2005), defined by the interval between the highest intraspecific distances and the lowest interspecific distances (DeSalle and Goldstein 2019). Then, a threshold for species delimitation may be established to the target taxonomic rank (Hebert et al. 2003; Meyer and Paulay 2005). Initially, Hebert et al. (2004) proposed a standard threshold for animals: 10 times the mean intraspecific variation for the group under study, the "10-fold rule". Other values have been refined to particular taxa (e.g., Meyer and Paulay 2005; Prantoni et al. 2018), and also alternative methods were proposed to find thresholds (see Meier et al. 2006). Subsequent studies questioned the 10-fold rule (Frézal and Leblois 2008), mainly because of its weak biological background (Meyer and Paulay 2005). Conversely, empirical data for different nematode groups validated this method to set thresholds (e.g., Ferri et al. 2009; Derycke et al. 2010; Martínez-Arce et al. 2020).

The phylum Nematoda Rudolphi, 1808 is an abundant and speciose group among metazoans. Nematodes comprise around 25 000 valid species, with estimated diversity higher than 40 million species (Larsen et al. 2017). Nematodes occupy a wide variety of ecological niches, as both free-living and parasitic species (Blumenthal and Davis 2004). Nematode identification often relies on morphological characters—which may be subtle, subjective, dependant on other characters, show high phenotypic plasticity, or be featured only in a specific life stage or sex (Coomans 2002; Nadler 2002; Carneiro et al. 2017). For nematodes of medical and economic interest, an accurate taxonomic diagnosis is fundamental to understand transmission mechanisms, develop management strategies, and prevent the deleterious effects of parasitism (Jasmer et al. 2003; Ortiz et al. 2016).

The *cox1* gene has been successfully employed as a DNA barcode to identify nematodes (e.g., Elsasser et al. 2009; Ferri et al. 2009; Prosser et al. 2013), although its use is still incipient for this phylum. Most studies show success in molecular taxonomy of nematodes using other molecular markers, such as ribosomal regions ITS, 28S, and 18S, and mitochondrial genes *nad5* and *cytb* (Floyd et al. 2002; Bhadury et al. 2008; Armenteros et al. 2014; Qing et al. 2020). The diversity of nematode lineages included in DNA barcoding studies have increased over the last years, but the literature lacks a transversal work exploring a broader range of species (Abebe et al. 2011; Prosser et al. 2013).

Thus, we assessed *cox1* performance as a DNA barcode in seven nematode genera, seeking to (*i*) test efficiency based on barcode gap and Probability of Correct Identification (PCI) analyses, (*ii*) compare PCI between two

Table 1. Taxon sampling for the barcoding gap and species richness estimation analyses, including the number of valid species and the predominant lifecycle strategy for each genus.

		BOLD		GenBank		
Genus	Valid species	Sequences	Species labels	Sequences	Species labels	Lifecycle strategy
Anguillicola[a]	5	592	5	592	5	Animal parasite
Caenorhabditis[b]	40	109	15	108	16	Free-living
Heterodera[a]	80	346	26	931	29	Plant parasite
Meloidogyne[c,d]	122	343	28	367	27	Plant parasite
Onchocerca[e]	34	87	17	76	16	Animal parasite
Strongyloides[f]	38	734	7	787	8	Animal parasite
Trichinella[g]	9	102	9	67	9	Animal parasite
All datasets		2313	107	2928	110	

Note: References: a, Bezerra et al. 2020; b, caenorhabditis.org; c, Hunt and Handoo 2009; d, Álvarez-Ortega et al. 2019; e, Lefoulon et al. 2017; f, GBIF Secretariat 2019; g, Sharma et al. 2020.

public sequence databases, and (*iii*) estimate species richness in the compiled datasets through an automated distance-based species clustering tool. Issues related to operational biases and challenges in nematode taxonomy are discussed under the light of DNA barcoding.

Materials and methods

Data obtention and filtering

BOLD (boldsystems.org) and GenBank (ncbi.nlm.nih.gov/genbank) are open-access collections of annotated nucleotides. The former was developed to be a curated database of barcoding sequences (Ratnasingham and Hebert 2007), while the latter is a general repository for DNA sequences and their protein translations (Benson et al. 2018). In both databases, nucleotide sequences must be identified by the user to the lowest taxonomic level possible during submission. But only in BOLD, deposited sequences are double-checked, as BOLD administrators perform quality control of uploaded data (Ratnasingham and Hebert 2007). Thus, we expected that the PCI (see below) for this database would be higher when compared to GenBank.

Cox1 sequences were retrieved in October 2019 from GenBank and BOLD, generating separate datasets for seven nematode genera (see Results section and Table 1). The workflow for sequence acquisition, curation, and analysis was based on Kvist (2014) and Sundberg et al. (2016). We looked for conspicuous genera from distinct fields of science (e.g., medical, agricultural) that also represented the diversity of lifestyles within Nematoda (e.g., free-living, plant parasites, and animal parasites), giving priority to taxa with higher sequence availability in public databases. To ensure robust analyses, we excluded duplicates, unverified sequences, and those identified only to generic level (e.g., *Trichinella* sp.).

We restricted our analyses to the 658 bp barcoding region of *cox1*, as defined by The Consortium for the Barcode of Life (Ratnasingham and Hebert 2007). Sequences were aligned using MAFFT 7.0 (Katoh et al. 2019), enabling direction adjustment and keeping other parameters in default. The software AliView (Larsson 2014) was used to visualise the alignments and to verify the reading frame. As a final control step, sequences shorter than 300 bp were removed. The following analyses were performed for each dataset individually.

Barcoding gap and Probability of Correct Identification (PCI) analyses

PAUP* 4.0 (Swofford 2002) was used to estimate uncorrected p-distances, ignoring missing data to affected sites and considering equal substitution rates to variable sites. Uncorrected p-distances yield more accurate (or at least similar) results when compared to other models of nucleotide evolution (e.g., K2P; see Srivathsan and Meier 2012; Collins et al. 2012). Output values were sorted in inter- and intraspecific bins. We followed Badotti et al. (2017) to verify the barcoding gap, plotting in Microsoft Excel boxplots of both intra- and interspecific distances. When possible, a barcoding gap was delimited, considering the maximum intraspecific limit and the minimum interspecific limit assigned by the whiskers. For comparison purposes, a threshold for species delimitation based on the 10-fold rule (Hebert et al. 2004) was also calculated for each dataset.

The success of DNA barcoding in specimen identification does not rely on a well-defined threshold; it can be used for this purpose even when inter- and intraspecific distances overlap (see Collins and Cruickshank 2012). We calculated the PCI according to Hollingsworth et al. (2009) to evaluate the discriminative power of *cox1*. This analysis considered the maximum intraspecific distance and the minimum interspecific distance (or nearest-neighbour distance) for each species. If the maximum intraspecific distance of a species was less than the minimum interspecific distance, then specimen identification using *cox1* would be successful for that species (Hollingsworth et al. 2009). PCI values were presented as the percentage of species correctly identified. Singletons were excluded from this analysis, as it is not possible to calculate intraspecific distances in this case. We then followed the graphical approach suggested by Collins and Cruickshank (2012): PCI values were visualised in a

Fig. 1. Pairwise distance distribution of the seven analysed genera comprising sequences retrieved from BOLD. For each genus, intraspecific (dark blue) and interspecific (light blue) distances are shown. When possible, a barcoding gap was delimited considering the maximum intraspecific limit and the minimum interspecific limit assigned by the whiskers. The dots represent outlier values.

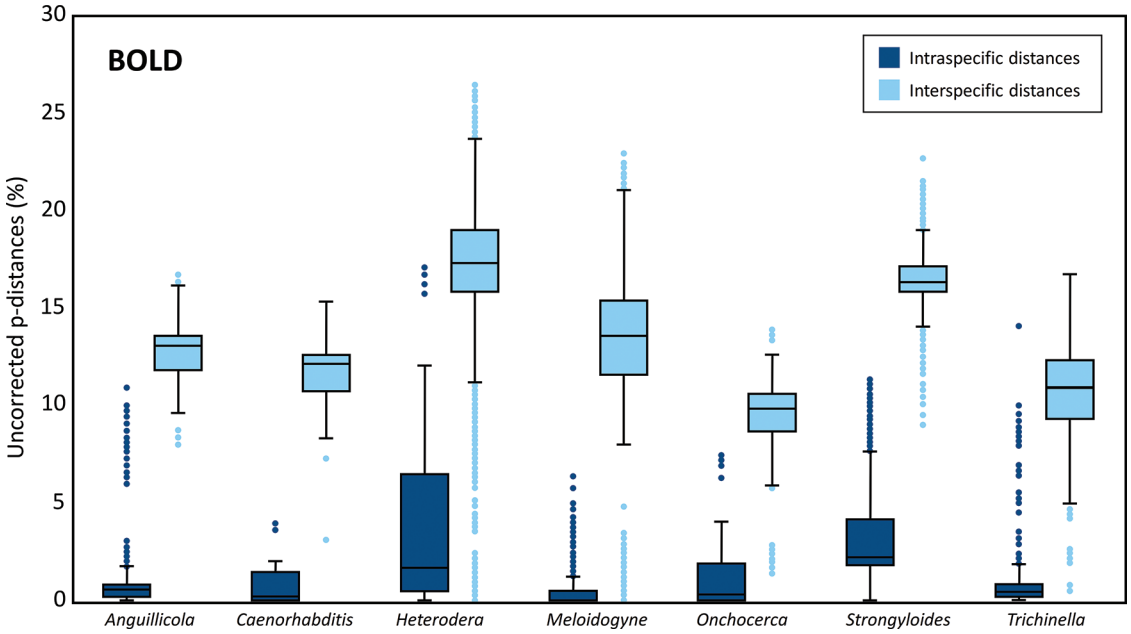
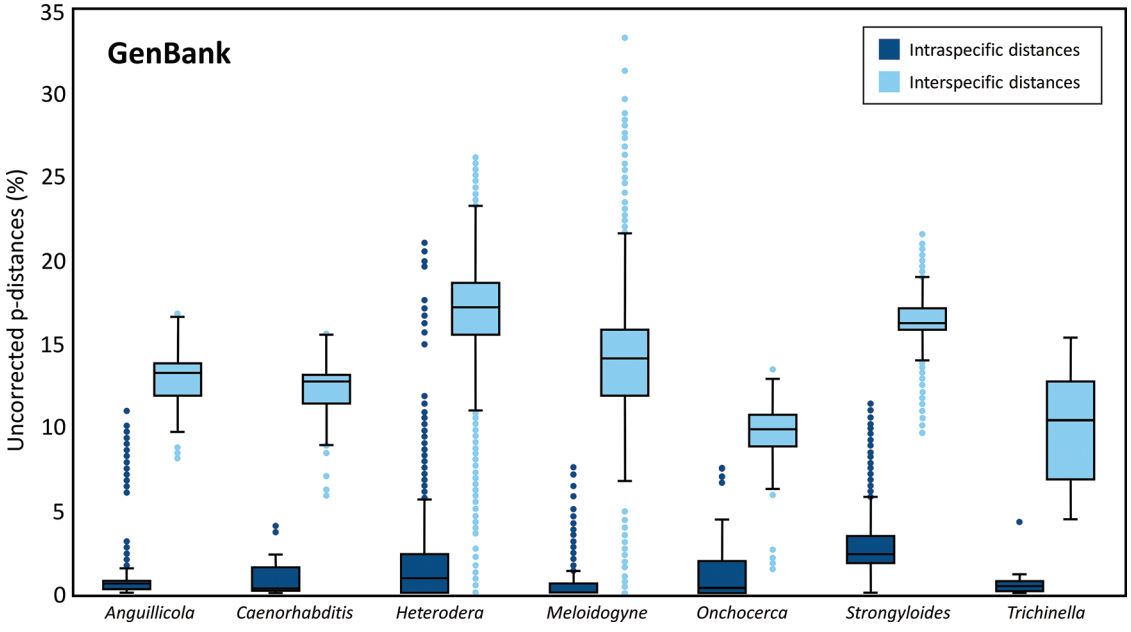


Fig. 2. Pairwise distance distribution of the seven analysed genera comprising sequences retrieved from GenBank. For each genus, intraspecific (dark blue) and interspecific (light blue) distances are shown. When possible, a barcoding gap was delimited considering the maximum intraspecific limit and the minimum interspecific limit assigned by the whiskers. The dots represent outlier values.



scatter plot, using a 1:1 reference slope to represent the point at which the difference between the two variables is zero. Finally, PCIs of BOLD and GenBank were converted to a 2 × 2 contingency table, and PAST 3.26 (Hammer et al. 2001) was used to perform Fisher's Exact Test and compare identification success between databases.

Species richness estimation

To test the applicability of DNA barcoding in species discovery, hypotheses of species richness should be estimated ignoring taxonomic labels (Collins and Cruickshank 2012). We then used Automatic Barcode Gap Discovery (ABGD; Puillandre et al. 2012) to estimate the richness

Table 2. Results from the barcoding gap analyses of sequences retrieved from BOLD and GenBank, showing in percent mean intraspecific distance, gap range visualised in boxplots, and threshold values calculated following the 10-fold rule.

	BOLD			GenBank		
Genus	Mean intraspecific distance	Gap range	Threshold sensu 10-fold rule	Mean intraspecific distance	Gap range	Threshold sensu 10-fold rule
Anguillicola	0.57	1.52–9.63	5.70	0.57	1.45–9.63	5.70
Caenorhabditis	0.76	1.96–8.30	7.60	0.79	2.28–8.87	7.90
Heterodera	3.47	NC	34.70	2.30	5.60–10.90	23.00
Meloidogyne	0.46	1.26–8.02	4.60	0.43	1.27–6.66	4.30
Onchocerca	1.60	4.03–5.92	16.00	1.71	4.38–6.19	17.10
Strongyloides	2.87	7.68–14.10	28.70	2.80	5.75–13.89	28.00
Trichinella	1.13	1.84–4.95	11.30	0.44	1.11–4.37	4.40

Note: NC, not computed—when whisker overlapping did not allow a barcode gap detection.

(number of species) in the obtained datasets. This tool clusters sequences in hypothetical species based on the statistical inference of a barcoding gap; the results are then used in a recursive analysis (Puillandre et al. 2012). We ran the analyses at the web interface (bioinfo.mnhn.fr/abi/public/abgd) using default parameters (P_{min} = 0.001; P_{max} = 0.1; Nb bins = 10; X = 1.5) and simple distance.

ABGD generates several hypotheses when used to estimate richness, and choosing the most reliable hypothesis can be a challenging task (Kekkonen and Hebert 2014). We interpreted ABGD results using a prior intraspecific divergence limit of P = 0.01 because it reproduces with higher correspondence the practical delimitations made by taxonomists, emphasising stringency to avoid theoretical overestimation (Puillandre et al. 2012). Thus, ABGD estimates were compared to the number of species labels informed by BOLD and GenBank.

Results

Database compilation

A total of 5241 *cox1* sequences composed our datasets. From BOLD, 2313 sequences were retrieved, representing 107 species labels (species names); the other 2928 sequences were retrieved from GenBank, representing 111 species labels (Table 1). Our dataset comprises organisms with distinct lifestyles (Table 1) that are historically independent and discordant with taxonomic classifications (Coomans 2002). We followed the classification proposed by De Ley and Blaxter (2004), as it is currently the most comprehensive taxonomic system for Nematoda and it matches the NCBI taxonomy database (Sayers et al. 2009; Benson et al. 2018). However, we are aware that hierarchy above family level may vary among authors and research groups (e.g., Eyualem et al. 2006). Our analyses approached seven datasets for each database, comprising the following genera: *Anguillicola* Yamaguti, 1935 (Spirurina: Anguillicolidae); *Caenorhabditis* Osche, 1952 (Rhabditina: Rhabditidae); *Heterodera* Schmidt, 1871 (Tylenchina: Heteroderidae); *Meloidogyne* Goeldi, 1892 (Tylenchina: Meloidogynidae); *Onchocerca* Diesing, 1841 (Spirurina: Onchocercidae); *Strongyloides* Grassi, 1879 (Tylenchina: Strongyloididae); and *Trichinella* Railliet, 1895 (Trichinellida: Trichinellidae). For a list of all sampled species, see Table S1[1].

Barcoding gap

Based on the pairwise distance boxplots for the analysed taxa (Figs. 1 and 2), we categorised *cox1* efficiency into the three categories proposed by Badotti et al. (2017): "good", "intermediate", and "poor". We considered the *cox1* efficiency good for genera that presented a clear gap between intra- and interspecific distances, even if outliers overlapped; intermediate when the whiskers of intra- and interspecific distances overlapped; and poor whenever the boxes overlapped.

Cox1 showed a conspicuous barcoding gap in most of the tested datasets. The sequences from BOLD showed good efficiency of *cox1* for six out of seven analysed genera: *Anguillicola*, *Caenorhabditis*, *Meloidogyne*, *Onchocerca*, *Strongyloides*, and *Trichinella*; and intermediate efficiency for *Heterodera* (Fig. 1). The efficiency of *cox1* was good for all genera retrieved from GenBank (Fig. 2). Outlier values were present in all pairs of intra- and interspecific comparisons for both databases.

We emphasise the significant variation of the barcoding gap among investigated genera, ranging from a low of 1.1% (*Trichinella*, GenBank) to a high of 14.1% (*Strongyloides*, BOLD) (Table 2). Moreover, the gap values found in the boxplots are incongruent with the thresholds obtained through the 10-fold rule (Table 2). Most threshold values obtained via the 10-fold rule would fail in the empirical discrimination of species, especially to *Heterodera* and *Strongyloides*, which sequences exhibit an unusually high mean intraspecific distance in both datasets.

Probability of Correct Identification (PCI)

The overall PCIs obtained for BOLD (72.72%) and GenBank (70.11%) (Fig. 3) were not statistically different (Fisher's Exact Test, p = 0.7399, Table 3). The PCI varied among genera (Table 3). The highest PCI values (100%) were detected for *Caenorhabditis* (both databases) and *Trichinella* (GenBank).

[1]Supplementary data are available with the article at https://doi.org/10.1139/gen-2020-0140.

Anguillicola and *Caenorhabditis* were the only genera that presented the same values for both databases (Table 3). However, the differences found between the PCIs of BOLD and GenBank were not significant for any genus (Table 3).

Species richness estimation

The species richness estimated by ABGD was different from the species numbers informed by the databases for any of the analysed genera (Table 4): it was higher than the number of taxonomic labels informed by BOLD and GenBank for most genera, including *Anguillicola*, *Caenorhabditis*, *Onchocerca*, *Strongyloides*, and *Trichinella*. The only exception was *Meloidogyne*, which showed a lower number of species than informed by the databases. Conversely, for *Heterodera* ABGD predicted a lower richness for BOLD data and a higher richness than expected for GenBank data.

Discussion

In this study, we explored *cox1* performance as a DNA barcode for different lineages of Nematoda, represented by seven genera. The barcoding gap analyses tested the applicability of this molecular marker in species discovery and delimitation. We found barcoding gaps for the seven analysed genera using GenBank sequences; for BOLD sequences, only six genera disclosed a barcoding gap. Moreover, we checked the hypothetical accuracy of the identifications (i.e., PCI), compared PCI between BOLD and GenBank, and estimated species richness based on *cox1* for each dataset (i.e., ABGD). We found PCI rates around 70% for both databases, and ABGD results overall pointed out to a higher species richness than the taxonomic labels informed by databases. These results highlight the prevalence of database issues and pitfalls in the widespread use of arbitrarily fixed species delimitation thresholds, the implications of which are relevant to a variety of metazoan lineages.

Barcoding gaps and fixed thresholds: a cautionary tale

The good performance of *cox1* for all the analysed genera in the barcoding gap analyses show the potential of this molecular marker as a tool to assess the diversity in Nematoda. The only exception was the intermediate performance for *Heterodera* sequences retrieved from BOLD. Accordingly, we recommend caution when defining divergence thresholds for species discovery.

Some authors have assigned fixed thresholds for nematode groups. Using the 10-fold rule, Ferri et al. (2009) estimated a 4.8% threshold for filarioid nematodes (also sampling *Onchocerca* species). For free-living marine nematodes, a 5% threshold obtained through the 10-fold rule is consistently being suggested to assess closely related and cryptic species of a "wide range of taxa" (Derycke et al. 2010; Armenteros et al. 2014; Martínez-Arce et al. 2020). Alternatively, a 2% threshold sorted out congeneric species from multiple lineages of parasites of vertebrates (Prosser et al. 2013). Moreover, previous works often

Fig. 3. Graphic summary of the Probability of Correct Identification (PCI) analyses of BOLD and GenBank sequences. For each sampled species, we compared the maximum intraspecific distance with the nearest-neighbour distance. A 1:1 slope represents the point at which the difference between the two variables is zero. In green (above the slope) species considered successfully identified; in red (below the slope) identification failures.

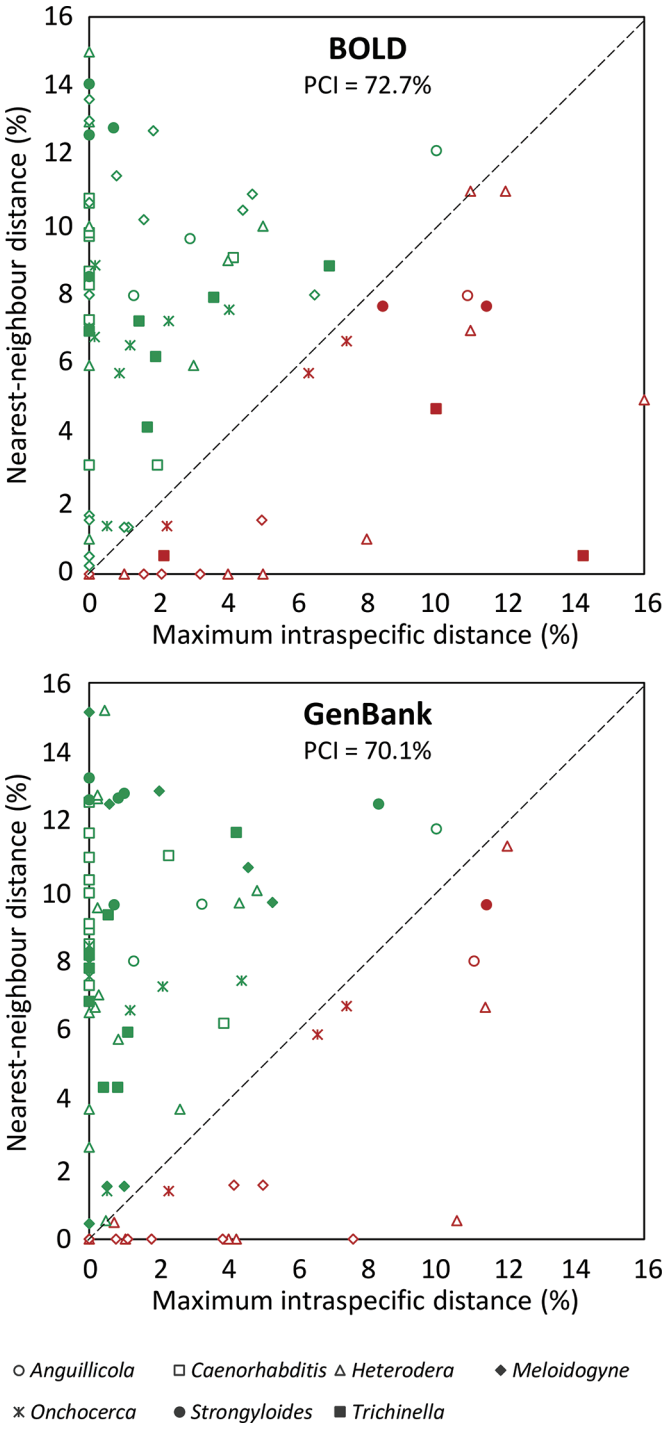


Table 3. Number of species labels (excluding singletons) and Probability of Correct Identification (PCI) for each of the seven genera sampled in BOLD and GenBank, with P-values (Fisher's Exact Test) for PCI comparison between databases.

	BOLD		GenBank		
Genus	Species labels	PCI (%)	Species labels	PCI (%)	*P*
Anguillicola	4	75.00	4	75.00	1.0000
Caenorhabditis	14	100.00	14	100.00	1.0000
Heterodera	22	54.54	24	58.33	1.0000
Meloidogyne	21	76.19	20	45.00	0.0578
Onchocerca	12	75.00	9	66.66	1.0000
Strongyloides	6	66.66	7	85.71	0.5594
Trichinella	9	66.66	9	100.00	0.2059
All datasets	88	72.72	87	70.11	0.7399

Table 4. Results from species richness estimation analyses, with number of species labels informed and richness estimated by Automatic Barcode Gap Discovery (ABGD) listed for BOLD and GenBank.

	BOLD		GenBank	
Genus	Species labels	ABGD richness	Species labels	ABGD richness
Anguillicola	5	9	5	9
Caenorhabditis	15	17	16	20
Heterodera	26	27	29	38
Meloidogyne	28	20	27	26
Onchocerca	17	18	16	18
Strongyloides	7	11	8	14
Trichinella	9	11	9	10

suggest a fixed threshold based on the lifecycle strategy of the scrutinised taxa (e.g., marine nematodes). We discourage this practice since the diversity of lifestyles within the phylum has emerged independently multiple times (Blaxter and Koutsovoulos 2015).

Here, we reiterate what Collins and Cruickshank (2012) postulated as "the sixth deadly sin of DNA barcoding": the inappropriate use of fixed thresholds for higher taxonomic levels. This assumption disregards the likely evolutionary heterogeneity and coalescence within diverse lineages (Fujita et al. 2012; Pentinsaari et al. 2016). A threshold value should be optimised from libraries of specific taxonomic groups (e.g., genus), putting away arbitrary "magic values" of divergence for higher taxonomic levels. Hence, an advantageous feature of DNA barcoding is its retroactive essence: as the accuracy of DNA barcoding upgrades the detection of taxa, it reciprocally enhances the correct labelling of library data.

We recognise that *cox1* performance in our dataset is far from flawless. The result for *Heterodera* (BOLD) was intermediate (Fig. 1), and the barcode gap analyses showed a remarkable number of outlier values in the boxplots (Figs. 1 and 2). Those outliers may be specimens from subsampled populations that present molecular distances above the conspecific average. However, we need to stress the likelihood of cryptic diversity and operational biases affecting the identification accuracy, as discussed below.

The position of the barcoding gap fluctuated among sampled genera (Table 2). This pattern emphasises their distinct coalescent times (Fujita et al. 2012) and an intrinsic divergence in *cox1* mutation rates among different lineages of Nematoda. It could be related (but not limited) to unique genome features (Molnar et al. 2011) or biological factors such as longevity (Cordero and Janzen 2013), asexuality (Lunt 2008), population size (Estes et al. 2004), generation time (Thomas et al. 2010), and host mobility (Blouin et al. 1995). The intraspecific divergence ranges for each genus also reflects the genetic structure of nematode populations (Blouin et al. 1995; Cole and Viney 2018), which should be analysed individually. For instance, *Trichinella* is reportedly characterised by low intraspecific divergences (Cole and Viney 2018).

So, how should the barcoding gap be established? With caveats and carefulness. Many methods and techniques are premised on a comprehensive sampling that would include all populations and species from a lineage (Lim et al. 2012). As the detection of a barcoding gap is sensitive to the number of species (Meier et al. 2008) and specimens sampled (Fontaneto et al. 2015), then the analyses should be reviewed regularly whenever new samples are generated and deposited in the databases (see Qing et al. 2020). Presumably, the genetic diversity within a taxon should reach an asymptote as the heterogeneity within a lineage increase. This knowledge may facilitate the establishment of a more robust barcoding threshold for specific taxa. Exploratory studies must avoid a priori threshold values. In cases where there is a lack of data for a specific taxon, we suggest the careful use of the threshold from the closest lineage as possible.

In BOLD and GenBank we trust (but not blindly)

The power of DNA barcoding as a taxonomic tool is not limited to a global barcoding gap. Intra- and interspecific distances may overlap without invalidating the identification success (for a discussion, see Collins and Cruickshank 2012). The efficacy of a molecular marker in organism identification must then be evaluated on its own—this is the underlying idea of the PCI analysis (Hollingsworth et al. 2009; Badotti et al. 2017). Using *cox1* as a molecular barcode, we found a PCI around 70% for both BOLD and GenBank (Fig. 3; Table 3). In an ideal scenario, these rates would be closer to 100%, as reported for other metazoan groups (e.g., Blagoev et al. 2009; Pérez-Asso et al. 2016; Bakhoum et al. 2018).

Our PCI analyses reinforce the barcoding gap results (Figs. 1 and 2): the outlier comparisons exhibited on the histograms show an evident incongruence between the genetic distances and the database species labels. We aimed to test the performance of *cox1* to identify specimens. Considering the incipient application of this

marker for Nematoda, the observed PCI is remarkable. Here, improved species delimitation methods provide a step forward for future research that seeks to accurately assess diversity and identify specimens/sequences in different lineages of Nematoda.

The PCIs of BOLD and GenBank were statistically similar. These results dismiss our initial supposition that BOLD sequences would exhibit higher PCI compared to GenBank (see "Data obtention and filtering" section). Although BOLD mines sequences from GenBank, it is not a reason to assume that datasets (and the results obtained within) from both databases would be the same (e.g., Meiklejohn et al. 2019; Pentinsaari et al. 2020). Hence, GenBank sequences could equally contribute to exploratory analyses. Despite that, PCIs of most datasets concerning both databases are quite far from 100%, except for *Caenorhabditis*.

The use of data retrieved from any database—even curated ones—should not be done blindly. Careful screening may avoid unwanted problems. Indeed, misidentification and annotation errors are intrinsic from the way the sequences are deposited in public databases (e.g., Valkiūnas et al. 2008; Kvist 2014; Stavrou et al. 2018). The assignment errors may be related to operational biases, e.g., laboratory contamination, DNA of cells from the host, and data entry mistakes (Mutanen et al. 2016; Leray et al. 2019) but mainly because of specimen misidentification (see below) (Valkiūnas et al. 2008). The free access of these sequences may propagate these errors, and lead to erroneous conclusions (Valkiūnas et al. 2008). As the number of taxonomists has been decreasing, cases of misidentified sequences may increase soon (see Janssen et al. 2017). However, independent research groups have worked continuously to improve identification accuracy for different taxa and molecular markers (e.g., Heller et al. 2014; O'Leary et al. 2016; Dunlap et al. 2018). Curated datasets and pipelines for molecular identification have also been developed for nematode groups (e.g., Macheriotou et al. 2019; Qing et al. 2020). Efforts like these are invaluable resources, mainly for taxa which taxonomy is historically ambiguous, like Nematoda.

Hidden diversity, but to what degree?

The species richness estimated by ABGD mismatch the number of species labels informed by BOLD and GenBank for all datasets. ABGD is considered a conservative approach and recent studies reported its tendency to lump sequences belonging to different species, and seldom split conspecific sequences (Pentinsaari et al. 2017; Gélin et al. 2017; Busschau et al. 2019). The conservative proposal of this algorithm is desirable here since our aim was not to make taxonomic decisions, but warily to shed light on cryptic diversity and prominently dubious species boundaries. Nevertheless, species richness of the analysed genera was usually greater than taxonomic labels informed. Some of our results stood out, showing discrepancies both for underestimation, e.g., *Meloidogyne*, and for overestimation, e.g., *Heterodera* (Table 4). Remarkably, our sample encompasses all *Anguillicola* and *Trichinella* species (Table 1), and for both genera and datasets, ABGD estimated a higher richness. These cases could be investigated in-depth, integrating multiple sources of evidence to unravel the taxonomy of these groups.

Understudied taxa, such as Nematoda, are more likely to present a vague understanding of what a species is (Hey 2001; Nadler 2002). It worsens as many groupings of Nematoda lack a proper phylogenetic hypothesis (see Negreiros et al. 2019; Qing and Bert 2019). This problem applies to not only higher taxonomic levels (e.g., family and genus) but also to the monophyly of species (Nadler 2002). Nematode taxonomy is complicated by a high number of cryptic species (which may lead to an inherent underestimation of species richness) (Blaxter 2016), considerable intraspecific variation in morphology (Carneiro et al. 2017; Lee et al. 2017; Nyaku et al. 2018), and convergent morphological evolution via adaptation to similar lifestyles (or vice versa) (Blaxter and Koutsovoulos 2015).

The use of *cox1* as a DNA barcode usually allows taxonomic resolution at population/species level, but the peculiarities of each lineage may hinder species diagnosis (Powers et al. 2018). In genera such as *Meloidogyne*, the ancient asexuality and the hybrid origin of species has led to reticular evolutionary patterns that hamper the delimitation of apomictic species (Lunt 2008; Janssen et al. 2016; Powers et al. 2018). Still, approaches based only on mitochondrial DNA may overlook most recent speciation events due to the time-lag between speciation and haplotype lineage sorting to reciprocal monophyly (Nadler 2002).

Overall, *cox1* is a relevant tool for integrative taxonomy of nematodes

Our multiple analyses using *cox1* show the suitability of this molecular marker to the scrutiny of Nematoda at the genus and species level. The availability of data limited the approach adopted here. Thus, any taxon sampling bias, somehow, depicts the current trendings and state of the art on nematode research. We are aware that the data available in GenBank and BOLD, and so the genera coverage here, represent only a fraction of this diverse phylum. However, the use of thousands of sequences in a transversal study approaching different lineages with distinct lifecycle strategies has no precedent among Nematoda, as far as we know.

Overall, the results point out a substantial number of specimen misidentification or dubious species delimitation. For taxa with many cryptic species, complex morphology, and complex life histories, such as Nematoda, the taxonomic impediments arise. Thus, systematics become weakened whenever a single approach (e.g., morphology, molecular, behaviour) is prioritised (Coomans 2002). The term integrative taxonomy, coined around 15 years ago (Dayrat 2005; Will et al. 2005), uses multiple lines of evidence to inform taxonomy and is

widespread in the literature (Padial et al. 2010) but rarely used for nematodes. New species descriptions are often based exclusively on morphological comparisons of type specimens (e.g., Phillips et al. 2016; Acosta et al. 2017; Pinheiro et al. 2018).

The integration of large-scale and consistent DNA sequencing with traditional taxonomic approaches naturally improves the discovery of biological diversity and identification of specimens (Moritz and Cicero 2004). The use of *cox1* as a metazoan barcode enriches the large public databases, such as BOLD and GenBank, making them scientifically valuable (Fontaneto et al. 2015; Andújar et al. 2018). However, the success of the DNA barcoding strategy requires the maintenance of a reference database that obeys rigorous taxonomic criteria at the moment of the deposit of sequences, especially concerning voucher data (Ekrem et al. 2007). The standardisation of a molecular marker allows reliable cross-comparison between studies and databases (Smith et al. 2009), boosting its use in, e.g., applied sciences. *Cox1* can also improve metabarcoding studies to access nematode communities as 18S is usually inaccurate to species level and may even underestimate the real diversity (Tang et al. 2012; Blaxter 2016; Treonis et al. 2018). For identification purposes, a growing body of evidence shows *cox1* outperforming ribosomal markers, including 18S (Guardone et al. 2013; Singh et al. 2013; Armenteros et al. 2014) and ITS (Blouin 2002; Keskin et al. 2015). When feasible, the use of multi-locus barcode approaches should be preferred as they increase identification success (Meiklejohn et al. 2019).

After all, *cox1* barcoding is neither the panacea nor the archenemy of nematode taxonomy. We encourage the use of multiple methods to increase the robustness of taxonomic decisions. *Cox1* has been used extensively for varied groups of organisms and different taxonomic purposes (e.g., Zimmermann et al. 2015; Almerón-Souza et al. 2018; Gibbs 2018). Without a reliable taxonomic identification, all research carried out in academic and applied branches of life sciences are virtually worthless (Kholia and Fraser-Jenkins 2011). Therefore, *cox1* is a relevant ally in nematode systematics and taxonomy, improving other methodologies, aiding in cryptic diversity detection, and shedding light on specimen identification. In other words, *cox1* as a DNA barcode may be useful to tackle this can of worms.

Conflicts of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

Acknowledgements

The authors would like to thank Dr. Eliane Fraga da Silveira, Dr. Juliana Cordeiro, Dr. Suzana Bencke Amato, and Dr. Victor Hugo Valiati who provided valuable comments during the development of this article. The paper was significantly improved after critical reading and feedback from Dr. Emília Welter Wendt, Dr. Roger Vila, and Dr. Xue Qing. We also thank an anonymous reviewer and *Genome* Associate Editor Dr. Ian Hogg whose comments improved this manuscript. We acknowledge Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the fellowship of L.T.G. and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) for the fellowship of F.M.B. (Finance Code 001).

References

Abdo, Z., and Golding, G.B. 2007. A Step Toward Barcoding Life: A Model-Based, Decision-Theoretic Method to Assign Genes to Preexisting Species Groups. Syst. Biol. **56**(1): 44–56. doi:10.1080/10635150601167005. PMID:17366136.

Abebe, E., Mekete, T., and Thomas, W.K. 2011. A critique of current methods in nematode taxonomy. Afr. J. Biotechnol. **10**: 312–323.

Acosta, A.A., González-Solís, D., and da Silva, R.J. 2017. *Spinitectus aguapeiensis* n. sp. (Nematoda: Cystidicolidae) from *Pimelodella avanhandavae* Eigenmann (Siluriformes: Heptapteridae) in the River Aguapeí, Upper Paraná River Basin, Brazil. Syst. Parasitol. **94**(6): 649–656. doi:10.1007/s11230-017-9729-1. PMID:28466113.

Almerón-Souza, F., Sperb, C., Castilho, C.L., Figueiredo, P.I.C.C., Gonçalves, L.T., Machado, R., et al. 2018. Molecular identification of shark meat from local markets in Southern Brazil based on DNA barcoding: evidence for mislabeling and trade of endangered species. Front. Genet. **9**: 138. doi:10.3389/fgene.2018.00138. PMID:29755504.

Álvarez-Ortega, S., Brito, J.A., and Subbotin, S.A. 2019. Multigene phylogeny of root-knot nematodes and molecular characterization of *Meloidogyne nataliei* Golden, Rose & Bird, 1981 (Nematoda: Tylenchida). Sci. Rep. **9**(1): 11788. doi:10.1038/s41598-019-48195-0. PMID:31409860.

Andújar, C., Arribas, P., Yu, D.W., Vogler, A.P., and Emerson, B.C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Mol. Ecol. **27**(20): 3968–3975. doi:10.1111/mec.14844. PMID:30129071.

Armenteros, M., Rojas-Corzo, A., Ruiz-Abierno, A., Derycke, S., Backeljau, T., and Decraemer, W. 2014. Systematics and DNA barcoding of free-living marine nematodes with emphasis on tropical desmodorids using nuclear SSU rDNA and mitochondrial COI sequences. Nematology, **16**(8): 979–989. doi:10.1163/15685411-00002824.

Badotti, F., de Oliveira, F.S., Garcia, C.F., Vaz, A.B.M., Fonseca, P.L.C., Nahum, L.A., et al. 2017. Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). BMC Microbiol. **17**(1): 42. doi:10.1186/s12866-017-0958-x. PMID:28228107.

Bakhoum, M.T., Sarr, M., Fall, A.G., Huber, K., Fall, M., Sembène, M., et al. 2018. DNA barcoding and molecular identification of field-collected *Culicoides* larvae in the Niayes area of Senegal. Parasit. Vectors, **11**(1): 615. doi:10.1186/s13071-018-3176-y. PMID:30509304.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K.D., and Sayers, E.W. 2018. GenBank. Nucleic Acids Res. **46**(D1): D41–D47. doi:10.1093/nar/gkx1094. PMID:29140468.

Bezerra, T.N., Decraemer, W., Eisendle-Flöckner, U., Hodda, M., Holovachov, O., Leduc, D., et al. 2020. Nemys: World Database of Nematodes. WORMS. [In press]. doi:10.14284/366.

Bhadury, P., Austen, M.C., Bilton, D.T., Lambshead, P.J.D., Rogers, A.D., and Smerdon, G.R. 2008. Evaluation of combined morphological and molecular techniques for marine nematode (*Terschellingia* spp.) identification. Mar. Biol. **154**(3): 509–518. doi:10.1007/s00227-008-0945-8.

Blagoev, G., Hebert, P., Adamowicz, S., and Robinson, E. 2009. Prospects for using DNA barcoding to identify spiders in species-rich genera. Zookeys, **16**: 27–46. doi:10.3897/zookeys.16.239.

Blaxter, M. 2016. Imagining Sisyphus happy: DNA barcoding and the unnamed majority. Philos. Trans. R. Soc. B Biol. Sci. **371**(1702): 20150329. doi:10.1098/rstb.2015.0329. PMID:27481781.

Blaxter, M., and Koutsovoulos, G. 2015. The evolution of parasitism in Nematoda. Parasitology, **142**(S1): S26–S39. doi:10.1017/S0031182014000791. PMID:24963797.

Blouin, M.S. 2002. Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. Int. J. Parasitol. **32**(5): 527–531. doi:10.1016/S0020-7519(01)00357-5. PMID:11943225.

Blouin, M.S., Yowell, C.A., Courtney, C.H., and Dame, J.B. 1995. Host movement and the genetic structure of populations of parasitic nematodes. Genetics, **141**: 1007–1014. doi:10.1093/genetics/141.3.1007. PMID:8582607.

Blumenthal, T., and Davis, R.E. 2004. Exploring nematode diversity. Nat. Genet. **36**(12): 1246–1247. doi:10.1038/ng1204-1246. PMID:15565103.

Busschau, T., Conradie, W., and Daniels, S.R. 2019. Evidence for cryptic diversification in a rupicolous forest-dwelling gecko (Gekkonidae: *Afroedura pondolia*) from a biodiversity hotspot. Mol. Phylogenet. Evol. **139**: 106549. doi:10.1016/j.ympev.2019.106549. PMID:31265890.

Carneiro, R.M.D.G., F.S. de O, L., and Correia, V.R. 2017. Methods and Tools Currently Used for the Identification of Plant Parasitic Nematodes. *In* Nematology — Concepts, Diagnosis and Control. InTech. doi:10.5772/intechopen.69403.

Clare, E.L., Lim, B.K., Engstrom, M.D., Eger, J.L., and Hebert, P.D.N. 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol. Ecol. Notes. **7**(2): 184–190. doi:10.1111/j.1471-8286.2006.01657.x.

Cole, R., and Viney, M. 2018. The population genetics of parasitic nematodes of wild animals. Parasit. Vectors. **11**(1): 590. doi:10.1186/s13071-018-3137-5. PMID:30424774.

Collins, R.A., and Cruickshank, R.H. 2012. The seven deadly sins of DNA barcoding. Mol. Ecol. Resour. **13**: 969–975. doi:10.1111/1755-0998.12046. PMID:23280099.

Collins, R.A., Boykin, L.M., Cruickshank, R.H., and Armstrong, K.F. 2012. Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. Methods Ecol. Evol. **3**(3): 457–465. doi:10.1111/j.2041-210X.2011.00176.x.

Coomans, A. 2002. Present status and future of nematode systematics. Nematology, **4**(5): 573–582. doi:10.1163/15685410260438836.

Cordero, G.A., and Janzen, F. 2013. Does life history affect molecular evolutionary rates? Nat. Educ. **4**: 1.

Dayrat, B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. **85**(3): 407–415. doi:10.1111/j.1095-8312.2005.00503.x.

De Ley, P., and Blaxter, M. 2004. A new system for Nematoda: combining morphological characters with molecular trees, and translating clades into ranks and taxa. In Nematology Monographs and Perspectives 2. *Edited* by R. Cook and D.J. Hunt. Brill, Leiden, Netherlands. pp. 633–653.

Derycke, S., Vanaverbeke, J., Rigaux, A., Backeljau, T., and Moens, T. 2010. Exploring the Use of Cytochrome Oxidase c Subunit 1 (COI) for DNA Barcoding of Free-Living Marine Nematodes. PLoS One, **5**(10): e13716. doi:10.1371/journal.pone.0013716. PMID:21060838.

DeSalle, R., Egan, M.G., and Siddall, M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos. Trans. R. Soc. B Biol. Sci. **360**(1462): 1905–1916. doi:10.1098/rstb.2005.1722. PMID:16214748.

DeSalle, R., and Goldstein, P. 2019. Review and Interpretation of Trends in DNA Barcoding. Front. Ecol. Evol. **7**: 302. doi:10.3389/fevo.2019.00302.

Dunlap, C.A., Ramirez, J.L., Mascarin, G.M., and Labeda, D.P. 2018. Entomopathogen ID: a curated sequence resource for entomopathogenic fungi. Antonie Van Leeuwenhoek, **111**(6): 897–904. doi:10.1007/s10482-017-0988-2. PMID:29170835.

Ekrem, T., Willassen, E., and Stur, E. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol. Phylogenet. Evol. **43**(2): 530–542. doi:10.1016/j.ympev.2006.11.021. PMID:17208018.

Elsasser, S.C., Floyd, R., Hebert, P.D.N., and Schulte-Hostedde, A.I. 2009. Species identification of North American guinea worms (Nematoda: *Dracunculus*) with DNA barcoding. Mol. Ecol. Resour. **9**(3): 707–712. doi:10.1111/j.1755-0998.2008.02393.x. PMID:21564728.

Estes, S., Phillips, P.C., Denver, D.R., Thomas, W.K., and Lynch, M. 2004. Mutation Accumulation in Populations of Varying Size: The Distribution of Mutational Effects for Fitness Correlates in *Caenorhabditis elegans*. Genetics, **166**(3): 1269–1279. doi:10.1534/genetics.166.3.1269. PMID:15082546.

Eyualem, A., Andrássy, I., and Traunspurger, W. 2006. Freshwater Nematodes: Ecology and Taxonomy. CABI, Wallingford, UK.

Ferri, E., Barbuto, M., Bain, O., Galimberti, A., Uni, S., Guerrero, R., et al. 2009. Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). Front. Zool. **6**(1): 1. doi:10.1186/1742-9994-6-1. PMID:19128479.

Floyd, R., Abebe, E., Papert, A., and Blaxter, M. 2002. Molecular barcodes for soil nematode identification. Mol. Ecol. **11**(4): 839–850. doi:10.1046/j.1365-294X.2002.01485.x. PMID:11972769.

Fontaneto, D., Flot, J.-F., and Tang, C.Q. 2015. Guidelines for DNA taxonomy, with a focus on the meiofauna. Mar. Biodivers. **45**(3): 433–451. doi:10.1007/s12526-015-0319-7.

Frézal, L., and Leblois, R. 2008. Four years of DNA barcoding: Current advances and prospects. Infect. Genet. Evol. **8**(5): 727–736. doi:10.1016/j.meegid.2008.05.005. PMID:18573351.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., and Moritz, C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. **27**(9): 480–488. doi:10.1016/j.tree.2012.04.012. PMID:22633974.

Gale, K.R., and Crampton, J.M. 1987. DNA probes for species identification of mosquitoes in the *Anopheles gambiae* complex. Med. Vet. Entomol. **1**(2): 127–136. doi:10.1111/j.1365-2915.1987.tb00333.x. PMID:2979526.

GBIF Secretariat. 2019. GBIF Backbone Taxonomy. doi:10.15468/39omei.

Gélin, P., Postaire, B., Fauvelot, C., and Magalon, H. 2017. Reevaluating species number, distribution and endemism of the coral genus *Pocillopora* Lamarck, 1816 using species delimitation methods and microsatellites. Mol. Phylogenet. Evol. **109**: 430–446. doi:10.1016/j.ympev.2017.01.018. PMID:28219759.

Gibbs, J. 2018. DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. Genome, **61**(1): 21–31. doi:10.1139/gen-2017-0096. PMID:28972864.

Goldstein, P.Z., Desalle, R., Amato, G., and Vogler, A.P. 2000. Conservation Genetics at the Species Boundary. Conserv. Biol. **14**(1): 120–131. doi:10.1046/j.1523-1739.2000.98122.x.

Guardone, L., Deplazes, P., Macchioni, F., Magi, M., and Mathis, A. 2013. Ribosomal and mitochondrial DNA analysis of Trichuridae nematodes of carnivores and small mammals. Vet. Parasitol. **197**(1–2): 364–369. doi:10.1016/j.vetpar.2013.06.022. PMID:23920054.

Hammer, Ø., Harper, D.A., and Ryan, P.D. 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and DeWaard, J.R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. London. Ser. B Biol. Sci. **270**(1512): 313–321. doi:10.1098/rspb.2002.2218. PMID:12614582.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., and Francis, C.M. 2004. Identification of Birds through DNA Barcodes. PLoS Biol. **2**(10): e312. doi:10.1371/journal.pbio.0020312. PMID:15455034.

Heller, P., Tripp, H.J., Turk-Kubo, K., and Zehr, J.P. 2014. ARBitrator: a software pipeline for on-demand retrieval of auto-curated nifH sequences from GenBank. Bioinformatics, **30**(20): 2883–2890. doi:10.1093/bioinformatics/btu417. PMID:24990605.

Hey, J. 2001. The mind of the species problem. Trends Ecol. Evol. **16**(7): 326–329. doi:10.1016/S0169-5347(01)02145-0. PMID:11403864.

Hey, J. 2009. Why should we care about species? Nat. Educ. **2**(5): 2.

Hollingsworth, M.L., Andra Clark, A., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., et al. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol. Ecol. Resour. **9**(2): 439–457. doi:10.1111/j.1755-0998.2008.02439.x. PMID:21564673.

Hunt, D.J., and Handoo, Z.A. 2009. Taxonomy, identification and principal species. *In* Root-knot Nematodes. *Edited* by R.N. Perry, M. Moens, and J.L. Starr. CABI, Cambridge, MA, U.S.A. pp. 55–97.

Janssen, T., Karssen, G., Verhaeven, M., Coyne, D., and Bert, W. 2016. Mitochondrial coding genome analysis of tropical root-knot nematodes (*Meloidogyne*) supports haplotype based diagnostics and reveals evidence of recent reticulate evolution. Sci. Rep. **6**(1): 22591. doi:10.1038/srep22591. PMID:26940543.

Janssen, T., Karssen, G., Couvreur, M., Waeyenberge, L., and Bert, W. 2017. The pitfalls of molecular species identification: a case study within the genus *Pratylenchus* (Nematoda: Pratylenchidae). Nematology, **19**(10): 1179–1199. doi:10.1163/15685411-00003117.

Jasmer, D.P., Goverse, A., and Smant, G. 2003. Parasitic Nematode Interactions With Mammals and Plants. Annu. Rev. Phytopathol. **41**(1): 245–270. doi:10.1146/annurev.phyto.41.052102.104023. PMID:14527330.

Katoh, K., Rozewicki, J., and Yamada, K.D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. **20**(4): 1160–1166. doi:10.1093/bib/bbx108. PMID:28968734.

Kekkonen, M., and Hebert, P.D.N. 2014. DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. Mol. Ecol. Resour. **14**(4): 706–715. doi:10.1111/1755-0998.12233. PMID:24479435.

Keskin, E., Koyuncu, C.E., and Genc, E. 2015. Molecular identification of *Hysterothylacium aduncum* specimens isolated from commercially important fish species of Eastern Mediterranean Sea using mtDNA cox1 and ITS rDNA gene sequences. Parasitol. Int. **64**(2): 222–228. doi:10.1016/j.parint.2014.12.008. PMID:25543079.

Kholia, B.S., and Fraser-Jenkins, C.R. 2011. Misidentification makes scientific publications worthless–save our taxonomy and taxonomists. Curr. Sci. **100**: 458–461.

Kloos, W.E., and Wolfshohl, J.F. 1982. Identification of *Staphylococcus* species with the API STAPH-IDENT system. J. Clin. Microbiol. **16**(3): 509–516. doi:10.1128/JCM.16.3.509-516.1982. PMID:6752190.

Kvist, S. 2014. Does a global DNA barcoding gap exist in Annelida? Mitochondrial DNA A DNA Mapp. Seq. Anal. **27**: 2241–2252. doi:10.3109/19401736.2014.984166. PMID:25431824.

Larsen, B.B., Miller, E.C., Rhodes, M.K., and Wiens, J.J. 2017. Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life. Q. Rev. Biol. **92**(3): 229–265. doi:10.1086/693564.

Larsson, A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics, **30**(22): 3276–3278. doi:10.1093/bioinformatics/btu531. PMID:25095880.

Lee, M.R., Canales-Aguirre, C.B., Nuñez, D., Pérez, K., Hernández, C.E., and Brante, A. 2017. The identification of sympatric cryptic free-living nematode species in the Antarctic intertidal. PLoS One, **12**(10): e0186140. doi:10.1371/journal.pone.0186140. PMID:28982192.

Lefoulon, E., Giannelli, A., Makepeace, B.L., Mutafchiev, Y., Townson, S., Uni, S., et al. 2017. Whence river blindness? The domestication of mammals and host-parasite co-evolution in the nematode genus *Onchocerca*. Int. J. Parasitol. **47**(8): 457–470. doi:10.1016/j.ijpara.2016.12.009. PMID:28344097.

Leray, M., Knowlton, N., Ho, S.-L., Nguyen, B.N., and Machida, R.J. 2019. GenBank is a reliable resource for 21st century biodiversity research. Proc. Natl. Acad. Sci. **116**(45): 22651–22656. doi:10.1073/pnas.1911714116. PMID:31636175.

Lim, G.S., Balke, M., and Meier, R. 2012. Determining Species Boundaries in a World Full of Rarity: Singletons. Species Delimitation Methods. Syst. Biol. **61**(1): 165–169. doi:10.1093/sysbio/syr030. PMID:21482553.

Lunt, D.H. 2008. Genetic tests of ancient asexuality in Root Knot Nematodes reveal recent hybrid origins. BMC Evol. Biol. **8**(1): 194. doi:10.1186/1471-2148-8-194. PMID:18606000.

Macheriotou, L., Guilini, K., Bezerra, T.N., Tytgat, B., Nguyen, D.T., Phuong Nguyen, T.X., et al. 2019. Metabarcoding free-living marine nematodes using curated 18S and CO1 reference sequence databases for species-level taxonomic assignments. Ecol. Evol. **9**(3): 1211–1226. doi:10.1002/ece3.4814. PMID:30805154.

Martínez-Arce, A., De Jesús-Navarrete, A., and Leasi, F. 2020. DNA Barcoding for Delimitation of Putative Mexican Marine Nematodes Species. Diversity, **12**(3): 107. doi:10.3390/d12030107.

Meier, R., Shiyang, K., Vaidya, G., and Ng, P.K.L. 2006. DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. Syst. Biol. **55**(5): 715–728. doi:10.1080/10635150600969864. PMID:17060194.

Meier, R., Zhang, G., and Ali, F. 2008. The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the Barcoding Gap" and Leads to Misidentification. Syst. Biol. **57**(5): 809–813. doi:10.1080/10635150802406343. PMID:18853366.

Meiklejohn, K.A., Damaso, N., and Robertson, J.M. 2019. Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. PLoS One, **14**(6): e0217084. doi:10.1371/journal.pone.0217084. PMID:31216285.

Meyer, C.P., and Paulay, G. 2005. DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biol. **3**(12): e422. doi:10.1371/journal.pbio.0030422. PMID:16336051.

Molnar, R.I., Bartelmes, G., Dinkelacker, I., Witte, H., and Sommer, R.J. 2011. Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*. Mol. Biol. Evol. **28**(8): 2317–2326. doi:10.1093/molbev/msr057. PMID:21368317.

Moritz, C., and Cicero, C. 2004. DNA Barcoding: Promise and Pitfalls. PLoS Biol. **2**(10): e354. doi:10.1371/journal.pbio.0020354. PMID:15486587.

Mutanen, M., Kivelä, S.M., Vos, R.A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., et al. 2016. Species-Level Para- and Polyphyly in DNA Barcode Gene Trees: Strong Operational Bias in European Lepidoptera. Syst. Biol. **65**(6): 1024–1040. doi:10.1093/sysbio/syw044. PMID:27288478.

Nadler, S. 2002. Species delimitation and nematode biodiversity: phylogenies rule. Nematology, **4**(5): 615–625. doi:10.1163/15685410260438908.

Negreiros, L.P., Tavares-Dias, M., Elisei, C., Tavares, L.E.R., and Pereira, F.B. 2019. First description of the male of *Philometroides acreanensis* and phylogenetic assessment of Philometridae (Nematoda: Dracunculoidea) suggest instability of some taxa. Parasitol. Int. **69**: 30–38. doi:10.1016/j.parint.2018.10.010. PMID:30389617.

Nyaku, S.T., Lutuf, H., and Cornelius, E. 2018. Morphometric Characterisation of Root-Knot Nematode Populations from Three Regions in Ghana. Plant Pathol. J. **34**: 544–554. doi:10.5423/PPJ.OA.05.2018.0081. PMID:30588227.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. **44**(D1): D733–D745. doi:10.1093/nar/gkv1189. PMID:26553804.

Ortiz, V., Phelan, S., and Mullins, E. 2016. A temporal assessment of nematode community structure and diversity in the rhizosphere of cisgenic *Phytophthora infestans*-resistant potatoes. BMC Ecol. **16**(1): 55. doi:10.1186/s12898-016-0109-5. PMID:27905931.

Padial, J.M., Miralles, A., De la Riva, I., and Vences, M. 2010. The integrative future of taxonomy. Front. Zool. **7**(1): 16. doi:10.1186/1742-9994-7-16. PMID:20500846.

Pentinsaari, M., Ratnasingham, S., Miller, S.E., and Hebert, P.D.N. 2020. BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? PLoS One, **15**(4): e0231814. doi:10.1371/journal.pone.0231814. PMID:32298363.

Pentinsaari, M., Salmela, H., Mutanen, M., and Roslin, T. 2016. Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. Sci. Rep. **6**(1): 35275. doi:10.1038/srep35275. PMID:27734964.

Pentinsaari, M., Vos, R., and Mutanen, M. 2017. Algorithmic single-locus species delimitation: effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. Mol. Ecol. Resour. **17**(3): 393–404. doi:10.1111/1755-0998.12557. PMID:27292571.

Pérez-Asso, A.R., Núñez-Aguila, R., and Genaro, J.A. 2016. Morphology and COI barcodes reveal four new species in the *lycieus* group of *Calisto* (Lepidoptera, Nymphalidae, Satyrinae.). Zootaxa, **4170**(3): 401. doi:10.11646/zootaxa.4170.3.1. PMID:27701235.

Phillips, G.C., Bernard, E.J., Pivar, R.K., Moulton, J.M., and Shelley, R. 2016. *Coronostoma claireae* n. sp. (Nematoda: Rhabditida: Oxyuridomorpha: Coronostomatidae) from the Indigenous Milliped *Narceus gordanus* (Chamberlain, 1943). J. Nematol. **48**(3): 159–169. doi:10.21307/jofnem-2017-023. PMID:27765989.

Pinheiro, R.H., da, S., Santana, R.L.S., Monks, S., Santos, J.N.D., and Giese, E.G. 2018. *Cucullanus marajoara* n. sp. (Nematoda: Cucullanidae), a parasite of *Colomesus psittacus* (Osteichthyes: Tetraodontiformes) in the Marajó. Brazil. Rev. Bras. Parasitol. Veterinária. **27**(4): 521–530. doi:10.1590/s1984-296120180072. PMID:30427526.

Powers, T., Harris, T., Higgins, R., Mullin, P., and Powers, K. 2018. Discovery and Identification of *Meloidogyne* Species Using COI DNA Barcoding. J. Nematol. **50**(3): 399–412. doi:10.21307/jofnem-2018-029. PMID:30451423.

Prantoni, A.L., Belmonte-Lopes, R., Lana, P.C., and Erséus, C. 2018. Genetic diversity of marine oligochaetous clitellates in selected areas of the South Atlantic as revealed by DNA barcoding. Invert. Syst. **32**(3): 524. doi:10.1071/IS17029.

Prosser, S.W.J., Velarde-Aguilar, M.G., León-Règagnon, V., and Hebert, P.D.N. 2013. Advancing nematode barcoding: A primer cocktail for the cytochrome c oxidase subunit I gene from vertebrate parasitic nematodes. Mol. Ecol. Resour. **13**: 1108–1115. doi:10.1111/1755-0998.12082. PMID:23433320.

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. **21**(8): 1864–1877. doi:10.1111/j.1365-294X.2011.05239.x. PMID:21883587.

Qing, X., and Bert, W. 2019. Family Tylenchidae (Nematoda): an overview and perspectives. Org. Divers. Evol. **19**(3): 391–408. doi:10.1007/s13127-019-00404-4.

Qing, X., Wang, M., Karssen, G., Bucki, P., Bert, W., and Braun-Miyara, S. 2020. PPNID: a reference database and molecular identification pipeline for plant-parasitic nematodes. Bioinformatics, **36**: 1052–1056. doi:10.1093/bioinformatics/btz707.

Ratnasingham, S., and Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System. Mol. Ecol. Notes. **7**(3): 355–364. doi:10.1111/j.1471-8286.2007.01678.x. PMID:18784790.

Rollinson, D., Walker, T., and Simpson, A. 1986. The application of recombinant DNA technology to problems of helminth identification. Parasitology, **92**(S1): S53–S71. doi:10.1017/S0031182000085693. PMID:3012444.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., et al. 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. **37**(Database): D5–D15. doi:10.1093/nar/gkn741. PMID:18940862.

Sharma, R., Thompson, P.C., Hoberg, E.P., Brad Scandrett, W., Konecsni, K., Harms, N.J., et al. 2020. Hiding in plain sight: discovery and phylogeography of a cryptic species of *Trichinella* (Nematoda: Trichinellidae) in wolverine (*Gulo gulo*). Int. J. Parasitol. **50**(4): 277–287. doi:10.1016/j.ijpara.2020.01.003. PMID:32171846.

Singh, N., Chaudhary, A., and Singh, H.S. 2013. Identification of two species of *Binema* Travassos, 1925 (Oxyurida: Travassosinematidae) based on morphological and sequence analysis of genomic (18S) and mitochondrial (Cox1) gene markers. J. Nematode Morphol. Syst. **16**: 173–180.

Smith, M.A., Fernandez-Triana, J., Roughley, R., and Hebert, P.D.N. 2009. DNA barcode accumulation curves for understudied taxa and areas. Mol. Ecol. Resour. **9**: 208–216. doi:10.1111/j.1755-0998.2009.02646.x. PMID:21564980.

Srivathsan, A., and Meier, R. 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics, **28**(2): 190–194. doi:10.1111/j.1096-0031.2011.00370.x.

Stavrou, A.A., Mixão, V., Boekhout, T., and Gabaldón, T. 2018. Misidentification of genome assemblies in public databases: The case of *Naumovozyma dairenensis* and proposal of a protocol to correct misidentifications. Yeast, **35**(6): 425–429. doi:10.1002/yea.3303. PMID:29320804.

Sundberg, P., Kvist, S., and Strand, M. 2016. Evaluating the Utility of Single-Locus DNA Barcoding for the Identification of Ribbon Worms (Phylum Nemertea). PLoS One, **11**(5): e0155541. doi:10.1371/journal.pone.0155541. PMID:27171471.

Swofford, D.L. 2002. PAUP* 4.0a167: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA, USA.

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., and Fontaneto, D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proc. Natl. Acad. Sci. **109**(40): 16208–16212. doi:10.1073/pnas.1209160109. PMID:22988084.

Thomas, J.A., Welch, J.J., Lanfear, R., and Bromham, L. 2010. A Generation Time Effect on the Rate of Molecular

Evolution in Invertebrates. Mol. Biol. Evol. **27**(5): 1173–1180. doi:10.1093/molbev/msq009. PMID:20083649.

Treonis, A.M., Unangst, S.K., Kepler, R.M., Buyer, J.S., Cavigelli, M.A., Mirsky, S.B., and Maul, J.E. 2018. Characterization of soil nematode communities in three cropping systems through morphological and DNA metabarcoding approaches. Sci. Rep. **8**(1): 2004. doi:10.1038/s41598-018-20366-5. PMID:29386563.

Valkiūnas, G., Atkinson, C.T., Bensch, S., Sehgal, R.N.M., and Ricklefs, R.E. 2008. Parasite misidentifications in GenBank: how to minimize their number? Trends Parasitol. **24**(6): 247–248. doi:10.1016/j.pt.2008.03.004. PMID:18440865.

Will, K.W., Mishler, B.D., and Wheeler, Q.D. 2005. The Perils of DNA Barcoding and the Need for Integrative Taxonomy. Syst. Biol. **54**(5): 844–851. doi:10.1080/10635150500354878. PMID:16243769.

Zimmermann, B.L., Campos-Filho, I.S., Deprá, M., and Araujo, P.B. 2015. Taxonomy and molecular phylogeny of the Neotropical genus *Atlantoscia* (Oniscidea, Philosciidae): DNA barcoding and description of two new species. Zool. J. Linn. Soc. **174**(4): 702–717. doi:10.1111/zoj.12256.