12-2018

Phylogenetic, Genomic and Morphological Investigations of Three Lance Nematode Species (*Hoplolaimus* spp.)

Xinyuan Ma
Clemson University, xm@g.clemson.edu

PHYLOGENETIC, GENOMIC AND MORPHOLOGICAL INVESTIGATIONS OF THREE LANCE NEMATODE SPECIES (*HOPLOLAIMUS* SPP.)

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Plant and Environmental Sciences

---

by
Xinyuan Ma
December 2018

Accepted by:
Dr. Paula Agudelo, Committee Chair
Dr. J. Antonio Baeza
Dr. William Bridges Jr.
Dr. Vincent Richards

ABSTRACT

Lance nematodes (*Hoplolaimus* spp.) are migratory ecto-endo plant-parasitic. They have been found from a wide range of the world that feed on the roots of a diversity of monocotyledonous and dicotyledonous plants, and have caused a great agricultural damage. Since more taxonomic knowledge and molecular references are demanded for the lance nematode phylogeny and population study, four chapters of lance nematode researches on three species were presented here: (1) A new species, *Hoplolaimus smokyensis* n. sp., was discovered from a mixed forest sample of maple (*Acer* sp.), hemlock (*Tsuga* sp.) and silverbell (*Halesia carolina*) from the Great Smoky Mountains National Park. It is characterized by possession of a lateral field with four incisures, an excretory pore posterior to the hemizonid, esophageal glands with three nuclei, phasmids anterior and posterior to the vulva, and the epiptygma absent. Phylogenetic analyses based on ribosomal and mitochondrial gene sequences also suggest *H. smokyensis* n. sp. to be an independent lineage distinct from all other reported *Hoplolaimus* species. (2) Additional morphological characteristics of *Hoplolaimus columbus* were described. Photos of its esophageal gland cell nuclei, a *H. columbus male* and abnormal female tails were presented. (3) The first complete *de novo* assembly of mitochondrial genome of *Hoplolaimus columbus* using Whole Genome Amplification and Illumina MiSeq technique was reported as a circularized DNA of 25228bp. The annotation results using two genetic codes were diagnosed and compared. Including *H. columbus*, phylogenetic relationships, gene content and gene order arrangement of 92 taxa nematodes were

analyzed. (4) The phylogenetic informativeness of mitochondrial genes in Nematoda Phylum is analyzed with two quantitative methods using mitochondrial genomes of 93 nematode species, including *H. columbus* and *H. galeatus*. Results from both methods agree with each other, indicate that the nad5 and nad4 contain higher informativeness than other candidates. Traditional markers like the cox1 and cytb genes contain medium informativeness. The nad4l and nad3 contain the lowest informativeness comparing with other protein-coding genes. Results also indicate that the phylogenetic informativeness is independent of the molecular sequence length of a phylogenetic marker. Concatenated-genes marker could present better phylogenetic informativeness if selected genes are higher informative.

DEDICATION

To life.

To my mother Jing Liu and father Tiejun Ma, who give me my precious life.

To my grandpa, all my kin and friends, who guide me to be an honest person.

To my future own family.

And

To love.

ACKNOWLEDGMENTS

I would like to thank Dr. Paula Agudelo with my highest gratitude, who helps and watches me to grow up/better during the past years. She is not only an excellent nematology scientist, but also a generous advisor with kindness and wisdom. I also give my sincerely appreciation to Dr. J. Antonio Baeza, Dr. William C. Bridges Jr. and Dr. Vincent Richards. They are all the best scientists I have ever met in my life, and they all inspired and taught me a lot.

I also want to thank Dr. Saara DeWalt, V Christine Minor and Department of Biological Sciences for offering me a Teaching Assistantship. I enjoyed helping students and sharing my knowledge with them. The teaching experience also helped myself to understand science better.

Moreover, I would like to thank Dr. Ernest C. Bernard from University of Tennessee, Dr. Robert Thomas Robbins from University of Arkansas and Dr. John Mueller from Clemson Edisto Research and Education Center for their wonderful nematology knowledge and technique support.

Thanks Mr. Baker, David, Jeanice, Wei, Juliet, Samara, Nathan, Brad, Claudia, and everyone in the Clemson Plant-parasitic Nematology Lab. They make me feel warm all the time, like a nice family.

Last but not least, I would like to thank Dr. Mohammad Rafiq Siddiqi for his careful and prudent work on nematode taxonomy, also thank all nematologists for their research. I'm so proud that I could contribute a little to nematology, and I will absolutely take full responsibility for all work in this dissertation.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF TABLES

List of Tables (Continued)

Table Page

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER I.

MORPHOLOGICAL AND MOLECULAR CHARACTERIZATION OF *HOPLOLAIMUS SMOKYENSIS* N. SP. (NEMATODA: HOPLOLAIMIDAE), A LANCE NEMATODE FROM THE GREAT SMOKY MOUNTAINS NATIONAL PARK, USA

Introduction

During a survey of fauna of the Great Smoky Mountains National Park, along the Tennessee–North Carolina border in the southeastern United States, a *Hoplolaimus* species was isolated from a mixed forest sample of maple (*Acer* sp.), hemlock (*Tsuga* sp.) and silverbell (*Halesia carolina*). This species was initially distinguished from other reported lance nematodes based on a BLAST (Basic Local Alignment Search Tool) search result of internal transcribed spacer 1 (ITS1) sequences at the NCBI (National Center for Biotechnology Information). Its position was isolated on a phylogenetic tree which was composed of all available unique ITS1 sequences of *Hoplolaimus* species from GenBank. A phylogenetic analysis of mitochondrial cytochrome oxidase subunit I (COI) sequences supported the former result, ensuring that molecular information of this undescribed lance nematode had not been reported yet. Females, males, and juveniles were examined for morphological characteristics, morphometrics, and phylogenetic relationships. We named the species *Hoplolaimus smokyensis* n. sp. to refer to the locality where it was first found.

To date, seven *Hoplolaimus* spp. have been reported from the southeastern USA, including *H. columbus* Sher, 1963, *H. concaudajuvencus* Golden & Minton, 1970, *H. galeatus* (Cobb, 1913) Thorne, 1935, *H. magnistylus* Robbins, 1982, *H. stephanus* Sher, 1963, *H. seinhorsti* Luc, 1958 and *H. tylenchiformis* von Daday, 1905 (Lewis and Fassuliotis, 1982). *Hoplolaimus columbus*, *H. galeatus* and *H. magnistylus* have been shown to be economically important and can cause serious damage to agronomic crops including cotton (*Gossypium hirsutum*), corn (*Zea mays*) and soybean (*Glycine max*) (Fassuliotis, 1974; Nyczepir and Lewis, 1979; Robbins *et al*., 1987, 1989; Henn & Dunn, 1989; Noe, 1993).

Morphologically, *H. smokyensis* n. sp. is closest to *H. galeatus* and *H. stephanus*, but it can be distinguished by minor morphological differences like the annuli in the lip region, morphometric values, and absence of epiptygma. In this study, we applied two genetic markers, mitochondrial DNA (COI) and ribosomal (ITS1), to obtain an analysis of higher resolution on taxonomic relationships, which support the delimitation of *H. smokyensis* n. sp.. DNA sequences were aligned by BLAST in GenBank and the highest similarity (89%) was with *H. magnistylus* sequences. Alignment analyses using MUSCLE (Edgar, 2004) show unique molecular characteristics possessed by *H. smokyensis* n. sp.. Phylogenetic trees generated by MrBayes (Huelsenbeck & Ronquist, 2001) indicate *H. smokyensis* n. sp. is a distinct clade.

Materials and methods

NEMATODE ISOLATION

Female, male, and juvenile specimens were sampled from the Great Smoky Mountains in July 2006. Nematodes were extracted from the soil by a combination of sieving-decanting and sucrose centrifugal-flotation. Specimens were killed and fixed, processed to glycerin, and permanently mounted on slides as described by Ye and Robbins (2003).

MORPHOLOGICAL OBSERVATION AND MICROGRAPHY

Permanently mounted specimens were examined using a Nikon Optiphot II compound microscope with Nomarski interference contrast. Using either a Nikon drawing tube or an ocular micrometer, measurements were made, which are expressed in micrometers (μm). Data is expressed as mean ± standard deviation with minimum to maximum range in parenthesis.

MOLECULAR PROFILES AND PHYLOGENY

DNA was extracted from individual nematodes using Sigma Extract-N-Amp kit (XNAT2) (Sigma, St. Louis, MO). The manufacturer's protocol was modified by reducing volumes to one eighth of the recommended amount (Ma, *et al.,* 2011). The Internal Transcribed Spacer (ITS1) was amplified with primers Hoc-1f (5'-AACCTGCTGCTGGATCATTA-3') and LSUD3r (5'-TATGCTTAAGTTCAGCGGGT-3') following Bae *et al.* (2009); and a portion of cytochrome c oxidase subunit I (COI)

using primers JB3 (5'-TTTTTTGGGCATCCTGAGGTTTAT-3') and JB5 (5'-AGCACCTAAACTTAAAACATAATGAAA-3') (Derycke *et al*., 2005). For COI, the initial denaturation was set at 95°C for 3min, followed by 33 cycles of 95°C for 45 s, 50°C for 1 min 15 s, 72°C for 2 min and final extension at 72°C for 10 min (Holguin *et al*., 2015). The amplified products were loaded onto a 1.5% agarose gel and visualized using GelRed™ (Biotium). PCR products for both regions were purified using magnetic beads and sequenced in both directions with the ABI 3730 capillary sequencer (Applied Biosystems) in the DNA Laboratory (School of Life Sciences) at Arizona State University. Sequencing results were edited and assembled in Sequencher 5.1 (Genes code corp., Ann Arbor, MI, USA). Consensus DNA sequences were searched in GenBank using BLAST, then aligned using MUSCLE (Edger, 2004).

A best-fit model of nucleotide substitution was selected using the GTR+I+G model with the Akaike Information Criterion (AIC) among 56 different models using ModelTest v 3.7 (Posada & Crandall, 1998**).** Bayesian inference was implemented for each gene separately using MrBayes 3.1.2 program (Huelsenbeck & Ronquist, 2001) running the chain for 10,000,000 generations with the Markov Chain Monte Carlo (MCMC) method, a sample frequency of 100 and burn-in value of 2500. We estimated the posterior probabilities of the phylogenetic trees (Larget & Simon, 1999) using the 50% majority rule. The phylogenetic trees were viewed on phylo.io (Robinson, 2016).

Results

Hoplolaimus smokyensis n. sp.

(Figs 1.1-1.3)

Measurements

See Table 1.1.

Description

Female

Female body generally cylindroid, vermiform, tapering slightly at ends. Midbody width *ca* 45μm (34.5-52.8μm). Head set off, with massive cephalic framework, usually bearing 6 lip annuli and the oral disc, occasionally 5 annuli (6 individuals out of 30 specimens). Labial region exhibiting sexual dimorphism, lower and more conical than the male's when viewed laterally, and circular *en face*. (Fig. 1.1A) The oral disc is surrounded by a lip annulus, which is separated into six sections -- two subdorsal, two subventral, and two reduced lateral sectors. Posteriorly, the head region is mostly tessellating. The basal lip annulus is subdivided into *ca* 18 equal blocks (Fig. 1.1B). Stylet long and robust, basal knobs tulip shaped, 48 μm (44.7-50.8μm) long. Dorsal esophageal gland orifice *ca* 12 μm from base of stylet knobs. Esophageal glands with three nuclei in total, only one nucleus was in the dorsal gland. Esophageal-intestinal junction extends 138μm (79.2-182.7μm) from anterior end. Esophageal lobes extend 198μm (168.5-223.3μm) from anterior end. Distinct nerve ring encircles isthmus.

Excretory pore prominent (Fig. 1.1 C), 163μm (144.1-217.2μm) from anterior end. Hemizonid large, about two annuli in length, located one or two annuli anterior to excretory pore (Fig. 1.1 D). Tail hemispherical to conoid-hemispherical, 24μm (20.3-34.5μm) in length (Fig. 1.1 E-F).

Lateral field incompletely areolated, has four lines through most of the body length, narrowing to two lines at the level of the metacorpus, ending near the level of the stylet base (Fig. 1.2A). Two phasmids large and conspicuous, variable in position. Right phasmid located anterior to vulva on two-thirds of the specimens. Vulva prominent, near midbody, with deep transverse silt. Epiptygma absent (Fig. 1.2B). Ovaries two, outstretched (Fig. 1.2C). Spermatocyte round to oval, usually with many sperm (Fig. 1.2D). Cuticular annulation at midbody distinct measuring 2 μm wide; subcuticular annulation distinct, about half as wide as outer cuticular annuli.

Male

Body shape similar to female, cylindroid, vermiform. Body length 1096μm (972.3-1260.7μm) , generally shorter than female. Head set off, labial region showing sexual dimorphism, higher and rounder than females when viewed laterally (Fig. 1.3A). Head region usually bearing 6 lip annuli and the oral disc, occasionally 5 annuli (4 individuals out of 17 specimens). Stylet length 43μm (40.6-48.7μm), stylet knob tulip-shape. Excretory pore located *ca* 156μm (136-172μm) from anterior end, 2-3 annuli posterior to hemizonid. Midbody width about 36μm (30.5-42.6μm). Cuticular annulation at midbody about 2 μm wide, distinct. Lateral field areolated, with four lines. Testis one,

outstretched anteriorly. Spicules 41μm (36.5-46.7μm), gubernaculum 10.5 μm (10-12μm). Bursa large and conspicuous (Fig. 1.3B).

Type material and locality

Specimens were obtained from the Great Smoky Mountains National Park, Sevier County, TN, USA, Laurel Falls Trail, at an elevation of 3307ft, in a mixed maple, hemlock and silverbell forest (GPS coordinates N 35° 40.874, W83° 36.149). Type specimens: Holotype female T-702t and another 5 Paratype slides (T-6864p – T-6868p containing 7 females and 7 males) are deposited in the Nematology Laboratory Collection, USDA, ARS, Beltsville, Maryland. Five other paratype slides (28506-28510), including 10 females, 7 males, and 5 juveniles, are deposited in Department of Nematology, University of California, Riverside.

Etymology:

The specific epithet refers to the locality where the species was first found.

Diagnosis and relationships

Hoplolaimus smokyensis n. sp. females have a straight body with, when relaxed with heat, a slightly ventrally inclined tail region. The labial region is characterized by 6 annuli, occasionally 5 annuli. The basal lip annulus subdivided with *ca* 24 longitudinal striae. Stylet length averages 47μm with robust tulip-shaped stylet knobs anterior projections. Hemizonid anterior to excretory pore *ca* 4μm. Lateral body with 4

continuous incisures start from metacorpus region to tail region, incompletely areolated. Lateral incisures on phasmid region irregularly areolated. Vulva epiptygma absent. Two scutella (phasmids) anterior and posterior to vulva. The male body length is shorter compared to the female. The male head region is higher and more rounded than the female when viewed laterally, possessing four lateral lines, prominent spicules and conspicuous bursa. Bursa extending to tail tip. Gubernaculum large, protrusible, with titillate and telamon.

Hoplolaimus smokyensis n. sp. differs from other reported 4-lateral-incisures species by: body length larger than *H. aorolaimoides* Siddiqi, 1972 (1.1-1.4 mm vs 0.8-0.92 mm); phasmid location in relation to vulva different from *H. californicus* Sher, 1963 (one anterior & one posterior vs both posterior); labial annuli more than *H. clarissimus* Fortuner, 1973 (6 vs 4), *H. galeatus* (6 vs 5), *H. stephanus* (6 vs 4), *H. tylenchiformis* Daday, 1905 (6 vs 3-4) and *H. sacchari* (Shamisi, 1979) Luc, 1981 (6 vs 3); juvenile tail shape rounded vs *H. concaudajuvencus* Golden & Minton, 1970 pointed; and stylet length and body length smaller than *H. magnistylus* Robbins, 1982 (43-49μm vs 52-61μm). Handoo & Golden published a key and diagnostic compendium including most of the 4-lateral-incisure species in 1992. Two years later, *H. igualaensis* Cid Del Prado V., 1994 was reported from Mexico, which had both scutella between the vulva and the anus.

Molecular profiles and phylogeny

Molecular sequences obtained in this study are deposited in the GenBank database with accession numbers KP230658, KP230659, KP303683, KP303684. All specimen information and GenBank accession numbers of *Hoplolaimus* species used in this study are listed in Table 1.2 (ITS1 sequences) and Table 1.3 (COI sequences). Sequences were first aligned using BLAST in GenBank. The results of both ITS1 and COI suggest *H. magnistylus* as the closest species to *H. smokyensis* n. sp. with 89% identity when query cover is 100%.

Sequences from all available lance nematode species with four lateral incisures were analyzed. Unique molecular characteristics of *H. smokyensis* n. sp. were marked out from alignment results using an asterisk symbol (Fig. 1.4-1.5). In the alignment of the ITS1 sequences and COI sequences, there are differences in 8 positions with the other five species.

Phylogenetic analyses (Fig. 1.6-1.7) based on the ribosomal and mitochondrial gene sequences suggest *Hoplolaimus smokyensis* n. sp. as a distinct lineage from *H. galeatus, H. stephanus, H. magnistylus* and *H. concaudajuvencus* (Bayesian posterior probability 100% in ITS1 tree and 100% in COI tree) and separate from other *Hoplolaimus* species like *H. columbus*. Comparing with other reported species, the phylogenetic results also suggested: i) *H. smokyensis* n. sp. sequences are classified into a clade which is parallel to other species. ii) *H. smokyensis* n. sp. was grouped with all other morphologically similar species in a big clade, but was separated into a different

clade from *H. columbus* or *H. seinhorsti*, which have 6 esophageal gland nuclei and only 1 lateral incisure.

Discussion

Hoplolaimus species are found feeding on the roots of a diversity of monocotyledonous and dicotyledonous plants. It has a wide distribution range in the United States (Wrather *et al.,* 1992; Lewis *et al.,* 1993; Martin *et al.,* 1994; Gazaway and McLean, 2003), and there are records from Canada, South America, Central America, and India on a variety of hosts (Fortuner, 1991). *H. tylenchiformis* was the first lance nematode collected by Daday (1905), which now is the type species for this genus. Based on differences in morphological characters, Siddiqi suggested dividing *Hoplolaimus* spp. into three subgenera according to lateral field incisures and the number of esophageal gland cell nuclei. The *Hoplolaimus* subgenus possesses a lateral field with four incisures and not obliterated; excretory pore normally closely behind hemizonid; and labial disc shape rounded. The other two subgenera, *Ethiolaimus* and *Basirolaimus*, have a lateral field with less than 4 lateral incisures and obliterated. The difference between *Ethilolaimus* and *Basirolaimus* is the number of esophageal gland nuclei, 3 for *Ethiolaimus* versus 6 for *Basirolaimus*. *Hoplolaimus smokyensis* n. sp. belongs to the *Hoplolaimus* subgenus.

In this study we have described *Hoplolaimus smokyensis* n. sp. from a mixed forest sample of maple, hemlock, and silverbell tree in the Great Smoky Mountains

National Park. Pathogenicity of *H. smokyensis* n. sp. in crop fields has not been reported yet. However, several lance nematode species, *H. stephanus*, *H. magnistylus* and *H. concaudajuvencus*, which have been reported to infect trees, are known to also feed on crops like corn, cotton, soybean, and turfgrass (Ma *et al.,* 2011, Holguin *et al.,* 2015). *H. galeatus* is also a prevalent pathogen of turf grasses such as St. Augustinegrass (*Stenotaphrum secundatum*) and bermudagrass (*Cynodon dactylon*) in Florida (Henn and Dunn, 1989; Giblin-Davis *et al*., 1990, 1995). These species are in the *Hoplolaimus* subgenus with similar morphological characteristics to *H. smokyensis* n. sp. Identification and description of this new species will contribute to studies of comparative biology and evolutionary biology of lance nematodes.

References

Bae, C.H., Szalanski, A.L. & Robbins, R.T. (2008). Molecular Analysis of the lance nematode, *Hoplolaimus* spp., using the first internal transcribed spacer and the D1-D3 expansion segments of 28S ribosomal DNA. *Journal of Nematology* 40, 201-209.

Bae, C.H., Robbins, R.T. & Szalanski, A.L. (2009). Molecular identification of some *Hoplolaimus* species from the USA based on duplex PCR, multiplex PCR and PCR-RFLP analysis. *Nematology* 11, 471-480.

Cid Del Prado V. I. (1994). Tres nuevos miembros de Hoplolaiminae (Nemata: Hoplolaimidae) de Mexico. *Nematropica* 24, 123-131.

Cobb, N.A. (1913). New nematode genera found inhabiting fresh water and non-brackish soils. *Journal of the Washington Academy of Sciences* 3, 432-444.

Daday, E.von. (1905). Untersuchungen über die Süsserwasser-Mikrofauna Paraguays. *Zoologica, Stuttgart* 18, 1-349.

Derycke, S., Remerie, T., Vierstraete, A., Backeljau, T., Vanfleteren, J., Vincx, M. & Moens, T. (2005). Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode Pellioditis marina. *Marine Ecology Progress Series* 300, 91-103.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792-1797.

Golden A.M. & Minton, N.A. (1969). Description and lavrval heteromorphism of *Hoplolaimus concaudajuvencus,* n. sp. (Nematoda: Hoplolaimidae). *Journal of Nematology* 2, 161-166.

Hall, T.A. (1999). Bioedit. A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.

Handoo, Z. & Golden, A.M. (1992). A key and diagnostic compendium to the species of the genus *Hoplolaimus* Daday, 1905 (Nematoda: Hoplolaimidae). *Journal of Nematology* 24, 45-53.

Holguin, C.M., Baeza, J.A., Mueller, J.D. & Agudelo, P. (2015). High genetic diversity and geographic subdivision of three lance nematode species (*Hoplolaimus* spp.) in the United States. *Ecology and Evolution* 5, 2929-2944.

Holguin, C.M., Ma, X., Mueller, J.D. & Agudelo, P. (2015). Distribution of *Hoplolaimus* species in soybean fields in South Carolina and North Carolina. *Plant Disease* 99, 1-5.

Huelsenbeck, J.P. & Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 1754-1755.

Larget, B. & Simon, D.L. (1999). Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* 16, 750-759.

Ma, X., Agudelo, P., Mueler, J.D. & Knap, H.T. (2011). Molecular characterization and phylogenetic analysis of *Hoplolaimus stephanus*. *Journal of Nematology* 43, 25-34.

Posada, D. & Crandall, K. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Robbins, R. T., Riggs, R. D. & Von Steen, D. (1987). Results of annual phytoparasitic nematode surveys of Arkansas soybean fields, 1978-1986. *Annals of Applied Nematology* 1: 50-55.

Robbins, R. T., Riggs, R. D. & Von Steen, D. (1989). Phytoparasitic nematode survey of Arkansas cotton fields, 1986-1988. *Supplement to the Journal of Nematology* 21: 619-623.

Robbins, R.T. (1982). Description of *Hoplolaimus magnistylus* n. sp. (Nematoda: Hoplolaimidae). *Journal of Nematology* 14, 500-506.

Robinson, O., Dylus, D., & Deesimoz, C. (2016). Phylo.io: Interactive viewing and comparison of large phylogenetic trees on the web. *Molecular Biology and Evolution* 33, 2163-2166.

Sher, S.A. (1963). Revision of The Hoplolaiminae (Nematoda) II. *Hoplolaimus* Daday, 1905 and *Aorolaimus* N. Gen. *Nematologica* 9, 267-296.

Siddiqi, M.R. (1972). Two new species of Scutellonema from cultivated soils in Africa with a description of *Hoplolaimus aorolaimoides* sp. n. from Portugal (Nematoda: Hoplolaiminae). *Proceedings of the Helminthological Society of Washington* 39, 7-13.

Siddiqi, M.R. (2000). *Tylenchida parasites of plants and insects*, *2nd edition*. Wallingford UK: CABI Publishing.

Ye, W., & Robbins, R.T. (2003). *Longidorus grandis* n. sp. and *L. paralongicaudatus* n. sp. (Nematoda: Longidoridae), two parthenogenetic species from Arkansas. *Journal of Nematology* 35, 375-387.

Table 1.1. Morphometric of *Hoplolaimus smokyensis* n. sp. holotype and paratypes. All measurements are in µm. Paratype values are means ± standard deviation with the range in parentheses.

Character	Female		Male
	Holotype	Paratypes	Paratypes
n	-	30	17
a	34.7	30.1±3.3 (23.5-36.9)	30.5±2.3 (26.1-34.5)
b	9.9	9.8±1.6 (7.1-14.9)	8.1±0.7 (7.1-9.3)
b'	7.3	6.8±0.9 (4.9-8.9)	5.8±0.5 (5.2-6.7)
c	115.7	114.5±15.8 (91.3-153.6)	36.8±3.7 (28.9-42.3)
c'	0.7	0.7±0.1 (0.5-1)	1.5±0.2 (1.3-1.9)
Lip annulus	6	5.8±0.4 (5-6)	5.8±0.4 (5-6)
Body length	1409	1334.9±174.2 (997-1870.5)	1096.4±86.7 (972.3-1260.7)
Lip height	8	8.1±0.3 (7.1-9.1)	8.1±0.2 (7.1-8.1)
Lip width	16	15.8±1 (14.2-18.3)	13.9±0.6 (12.2-14.2)
Cone	24	25.8±1.6 (22.3-28.4)	23.8±1.5 (22.3-26.4)
Stylet length	47	47.9±1.9 (44.7-50.8)	43.4±2.1 (40.6-48.7)
Esophageal length 1	142	138.1±16.7 (79.2-182.7)	136±6.4 (125.9-152.3)
Esophageal length 2	193	198.6±11.2 (168.5-223.3)	188.9±10.7 (172.6-207.1)
Body width	41	44.4±4.1 (34.5-52.8)	36.1±3.4 (30.5-42.6)
Tail length	26	24.9±4 (20.3-34.5)	30±3.4 (24.4-36.5)
Anal body width	37	35.5±3.1 (28.4-40.6)	20.2±1.3 (18.3-22.3)
Hemizonid from anterior end	162	157±14.2 (140.1-207.1)	150.3±8.8 (132-166.5)
Excretory pore from anterior end	166	163.6±14.7 (144.1-217.2)	156.1±8.6 (136-172.6)
Dorsal esophageal gland orifice	12	11.7±0.9 (10.2-12.2)	7±1 (6.1-8.1)
Vulva to anterior end distance	766	753.1±100.3 (601.5-1112.4)	-
V%	54.4	56.5±2.6 (49.7-62.1)	-
Spicule length	-	-	41.1±2.5 (36.5-46.7)
Gubernaculum	-	-	10.5±0.6 (10-12)

Table 1.2. Origin and GenBank accession numbers of *Hoplolaimus* species COI sequences used in this study.

Species	Location	Host	Host scientific name	Accession number
H. columbus	Barnwell County, SC	Sorghum bicolor	*Sorghum bicolor*	KP864583
H. columbus	Pierce County, GA	Soybean	*Glycine max*	KP864611
H. concaudajuvencus	Dallas County, TX	Bentgrass	*Agrostis* sp.	KP230667
H. concaudajuvencus	Dallas County, TX	Bentgrass	*Agrostis* sp.	KP230668
H. galeatus	St. Johns, FL	St. Augustine	*Stenotaphrum secundatum*	KP230564
H. galeatus	Baldwin County, SC	Bermuda	*Cynodon dactylon*	KP230554
H. magnistylus	Massac County, IL	Soybean	*Glycine max*	KP230588
H. magnistylus	Weakley County, TN	Corn	*Zea mays*	KP230657
H. stephanus	Riley County, KS	Bentgrass	*Agrostis* sp.	KP230593
H. stephanus	Warren County, OH	Bentgrass	*Agrostis* sp.	KP230626
H.smokyensis n. sp.	Sevier County, TN	Maple	*Acer* sp.	KP230658
H.smokyensis n. sp.	Sevier County, TN	Maple	*Acer* sp.	KP230659

Table 1.3. Origin and GenBank accession numbers of *Hoplolaimus* species ITS1 sequences used in this study.

Species	Location	Host	Host scientific name	Accession number
H. columbus	Pierce County, GA	Soybean	*Glycine max*	KP835333
H. columbus	Barnwell County, SC	Sorghum	*Sorghum bicolor*	KP835315
H. concaudajuvencus	Dallas County, TX	Bentgrass	*Agrostis* sp.	KP303685
H. concaudajuvencus	Dallas County, TX	Bentgrass	*Agrostis* sp.	KP303686
H. galeatus	Baldwin County, AL	Bermuda	*Cynodon dactylon*	KP303596
H. galeatus	St. Johns, FL	St. Augustine	*Stenotaphrum secundatum*	KP303607
H. magnistylus	Weakley County, TN	Corn	*Zea mays*	KP303681
H. magnistylus	Massac County, IL	Soybean	*Glycine max*	KP303634
H. seinhorsti	Alachua County, FL	Peanut	*Arachis hypogaea*	EU515327
H. seinhorsti	Fujian, China	-	-	KF486504
H. sp.1	Smoky Mountain, TN	-	-	EU515329
H. sp.2	Univeristy of Illinois, IS	Turfgrass	-	EU515330
H. sp.2	Manhattan, KS	Corn	*Zea mays*	EU515331
H. sp.3	Clemson, SC	Birch tree	*Betula* sp.	EU515332
H. sp.3	Limestone County, IA	Cotton	*Gossypium* sp.	EU515333
H. stephanus	Sargent County, ND	Soybean	*Glycine max*	KX347888
H. stephanus	Riley County, KS	Bentgrass	*Agrostis* sp.	KP303646
H. stephanus	Warren County, OH	Bentgrass	*Agrostis* sp.	KP303664
H.smokyensis n. sp.	Sevier County, TN	Maple	*Acer* sp.	KP303683
H.smokyensis n. sp.	Sevier County, TN	Maple	*Acer* sp.	KP303684

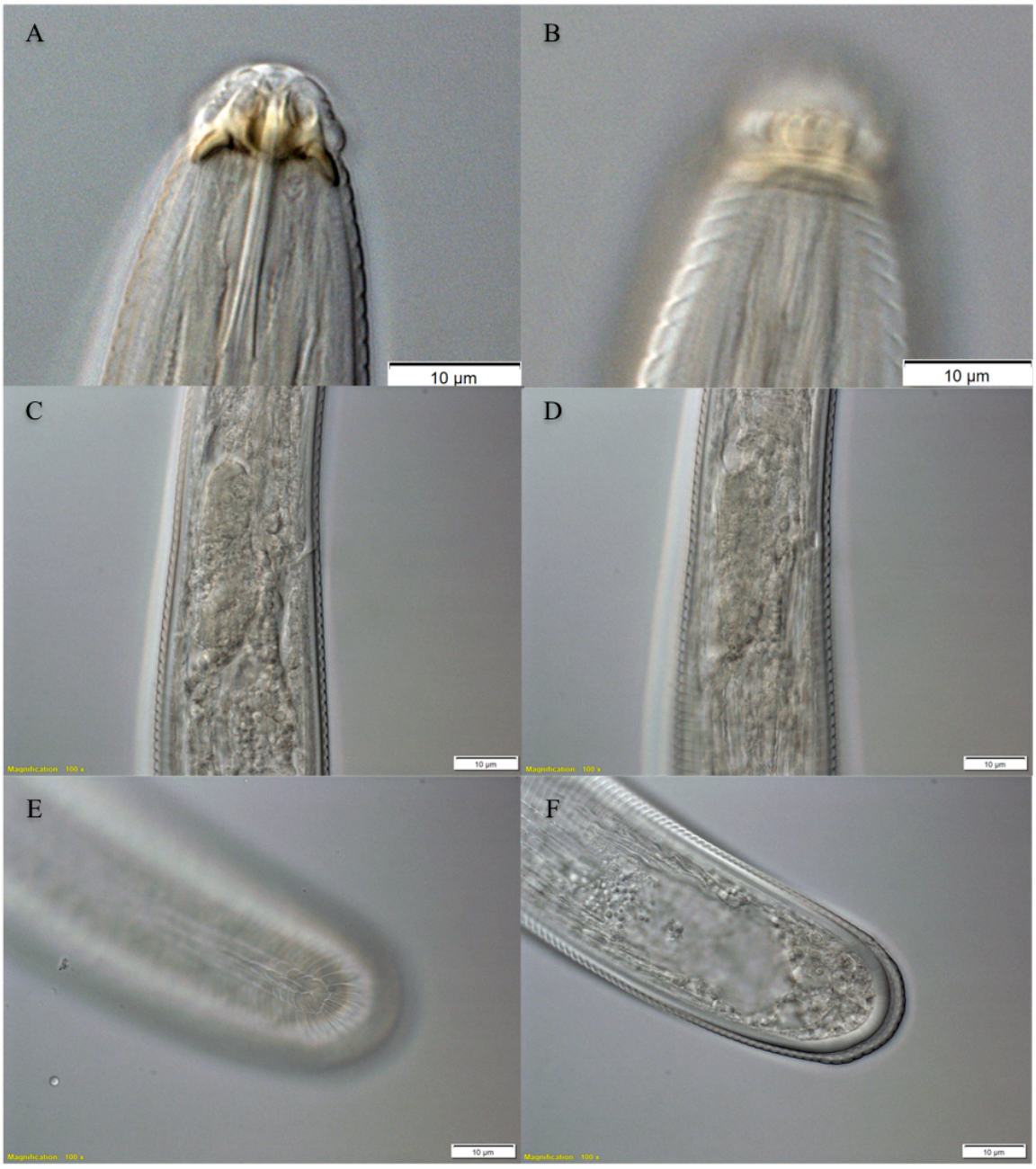

Fig. 1.1. Micrographs of *Hoplolaimus smokyensis* n. sp., females from the Great Smoky Mountains National Park. A-B: head of the same female specimen showing lip annuli and longitudinal striae.; C-D: esophageal region of the same female specimen showing excretory pore and hemizonid. E-F: tail of the same female specimen showing lateral field incisures, areolation and anus.

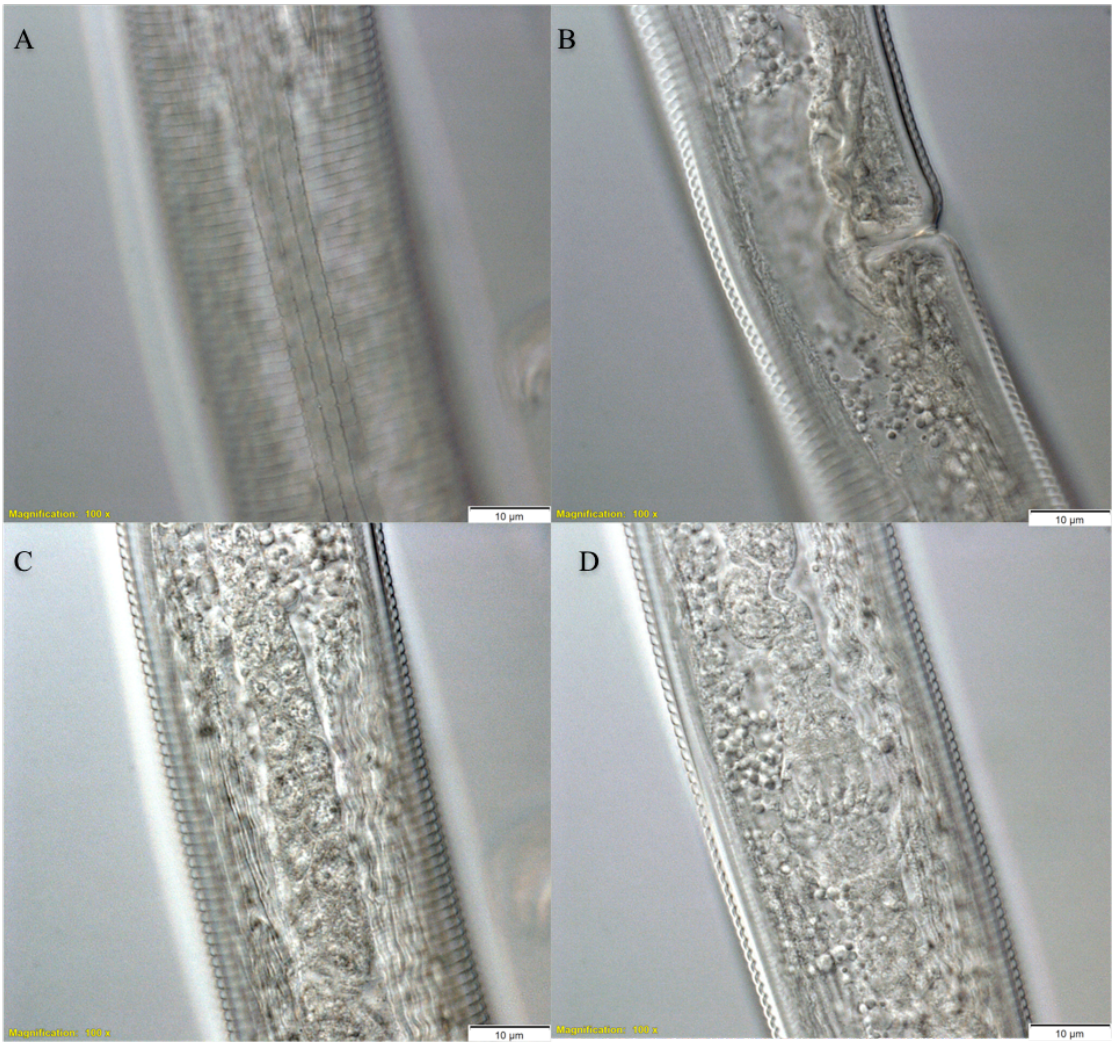

Fig. 1.2. Micrographs of *Hoplolaimus smokyensis* n. sp., females from the Great Smoky Mountains National Park. A: incomplete areolation in lateral field around phasmid; B: vulva region with absent epiptygma; C-D: ovary with sperm.

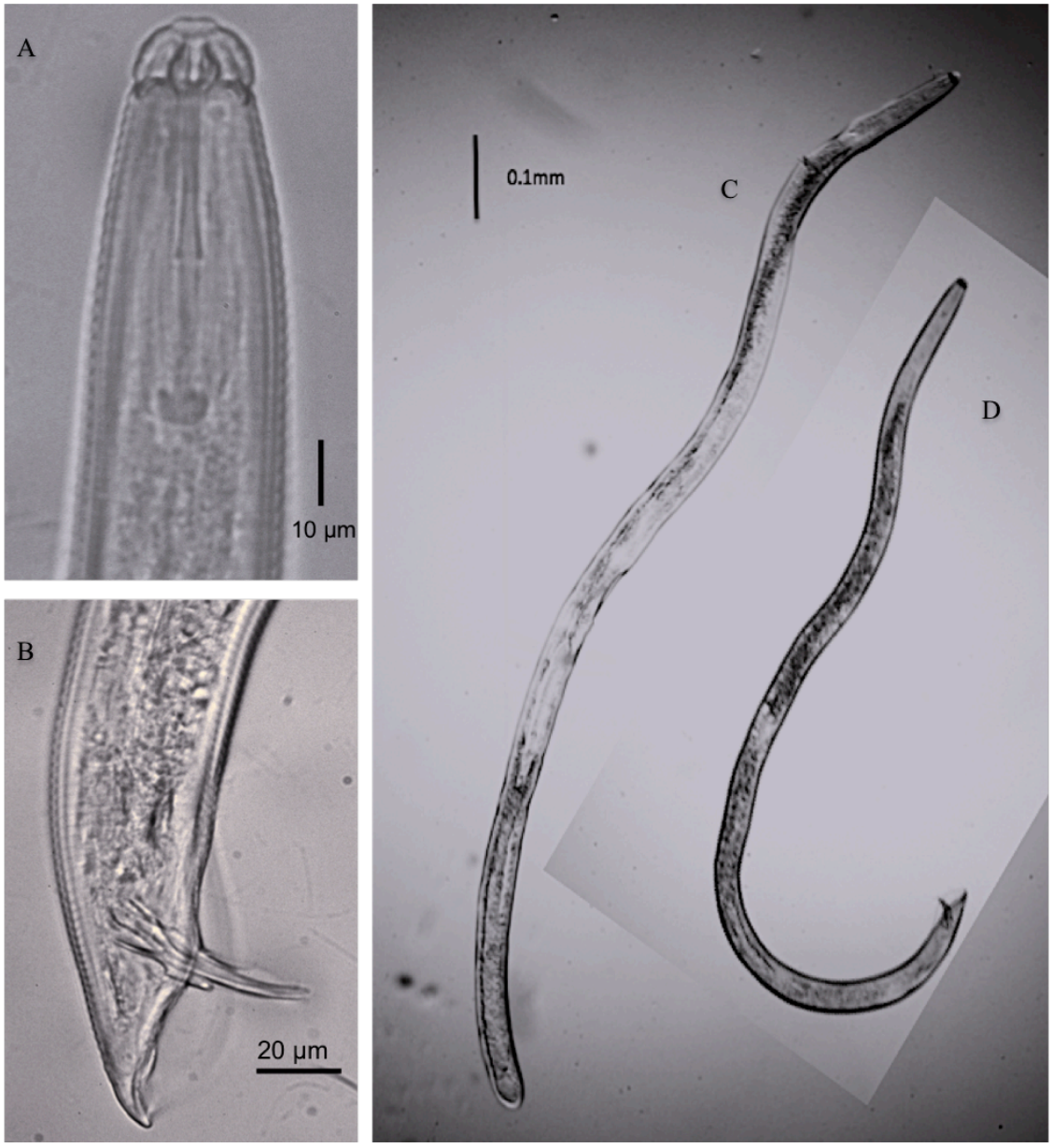

Fig. 1.3. Micrographs of *Hoplolaimus smokyensis* n. sp., from the Great Smoky Mountains National Park. A: male head; B: male tail; C-D: whole body micrographs of male and female under the same magnification.

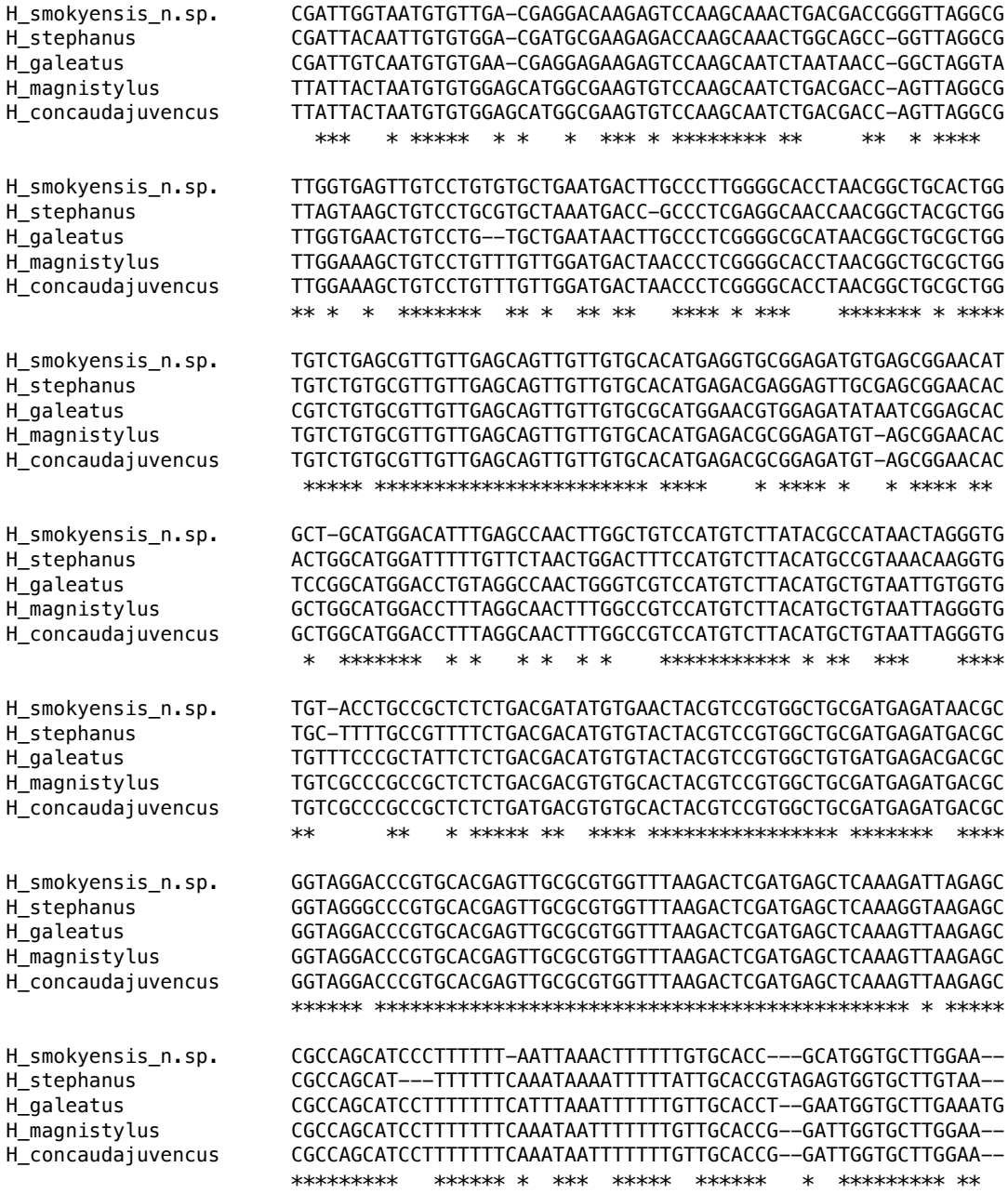

Fig. 1.4. Alignment of ITS1 sequences for *Hoplolaimus smokyensis* n. sp. and the four closest *Hoplolaimus* species. Positions without asterisk indicate polymorphisms, or differences between *Hoplolaimus smokyensis* n. sp. and at least one of the other four species.

```
H_smokyensis_n.sp. GGAAAAAAAAAATTGTTTGGTCATTTAGGGATAATTTATGCTATGCTTTCTATTGGGTTT
H_galeatus         GGAAAAAAAAAATTATTTGGACATTTGGGAATAATTTATGCGATAATTGCTATTGGTTTT
H_stephanus        GGAAAAAAGAAATTGTTTGGACATTTGGGGATAATTTATGCTATAATTTCTATTGGGTTT
H_magnistylus      GGAAAAAAAAAATTGTTTGGTCATTTAGGAATGATTTATGCTATAATTGCTATTGGTTTT
H_concaudajuvencus GGAAAAAAAAAATTATTTGGTCATTTAGGGTTGGTTTATGCTATGTTAGCTATTGGTTTT
                   ******** ***** ***** ***** **  *  ******* **  *  ******* ***

H_smokyensis_n.sp. ATTGGTTGTTTAGTTTGGGCTCATCATATATTTGTAGTTGGAATAGATTTAGATAGTCGA
H_galeatus         ATTGGGTGTTTGGTTTGGGCCCATCATATGTTTGTGGTTGGGATGGATTTGGATAGCCGG
H_stephanus        ATTGGTTGTTTGGTTTGAGCTCATCATATGTTTGTTGTGGGAATAGATTTGGATAGGCGG
H_magnistylus      ATTGGTTGTTTAGTTTGAGCTCATCATATGTTTGTTGTAGGAATAGACTTGGATAGACGT
H_concaudajuvencus ATTGGTTGTTTGGTTTGGGCTCATCATATATTTGTTGTTGGAATGGATTTAGATAGACGT
                   ***** ***** ***** ** ******** ***** ** ** ** ** ** ***** **

H_smokyensis_n.sp. GCTTATTTTAGTGCTGCTACAATGATTATCGCAGTACCTACTGGTATTAAAGTTTTTTCT
H_galeatus         GCTTATTTTAGGGCGGCTACTATAATTATTGCGGTTCCGACTGGGATTAAGGTTTTTTCC
H_stephanus        GCTTATTTTAGTGCAGCTTCTATAATTATTGCTGTTCCAACTGGTATTAAGGTTTTTTCT
H_magnistylus      GCTTATTTTAGGGCAGCAACGATGATTATTGCAGTTCCTACTGGTATTAAAGTTTTTTCT
H_concaudajuvencus GCTTATTTTAGTGCTGCTACTATAATTATTGCTGTTCCTACTGGTATTAAGGTTTTTTCT
                   ********** ** **  * ** ***** ** ** ** ***** ***** ********

H_smokyensis_n.sp. TGATTAATAACATTACATGCTTCTGTTTTATTTAATAGTTATTTATATGATTGGGTAATA
H_galeatus         TGGTTAATAACTTTGTTTTCTTCGATTTTTTTTAACAGTTACCTTTTGGATTGGGTGATA
H_stephanus        TGGTTAATAACTTTATATTCTTCGGTTTTGTTTGATAGATTTTTATTTGAATGAGTAATA
H_magnistylus      TGAATAATAACTTTATATGCTTCAGTTTTATTTAATAGTTATTTATATTATTGGGTTTTG
H_concaudajuvencus TGATTGATGACTTTATATTCTTCAATTTTATTTACTAGTTATTTATTTGAATGGGTAATA
                   **  * ** ** **   * ****  **** ***   ** *   * *   * ** **  *

H_smokyensis_n.sp. GGTTTTATTTATTTATTTACTATAGGGGGTTTAAGAGGTTTAGTTTTAAGTAATGCTAGT
H_galeatus         GGTTTTATTTATTTGTTTACAATTGGGGGTTTAAGTGGTTTAATTTTGAGGAACGCAAGT
H_stephanus        GGTTTTGTTTATTTATTTACTTTAGGTGGTTTGACTGGTTTGGTTTTGAGTAATGCTAGT
H_magnistylus      GGATTTATTTATTTGTTTACTTTGGGAGGTTTAAGTGGATTAATTTTAAGTAATGCTAGT
H_concaudajuvencus GGTTTTATTTATTTGTTTACTTTTGGTGGTTTAAGAGGTTTAGTTTTGAGAAATGCAAGT
                   ** *** ******* *****  * ** ***** *  ** **  **** ** ** ** ***

H_smokyensis_n.sp. TTGGATTTATTATTACATGATAC
H_galeatus         TTGGATTTGTTATTACATGACAC
H_stephanus        TTAGATTTATTGTTACATGATAC
H_magnistylus      TTAGATTTATTACTTCATGATAC
H_concaudajuvencus TTGGATTTATTATTACATGATAC
                   ** ***** **  * ***** **
```

Fig. 1.5. Alignment of COI sequences for *Hoplolaimus smokyensis* n. sp. and the four closest *Hoplolaimus* species. Positions without asterisk indicate polymorphisms, or differences between *Hoplolaimus smokyensis* n. sp. and at least one of the other four species.

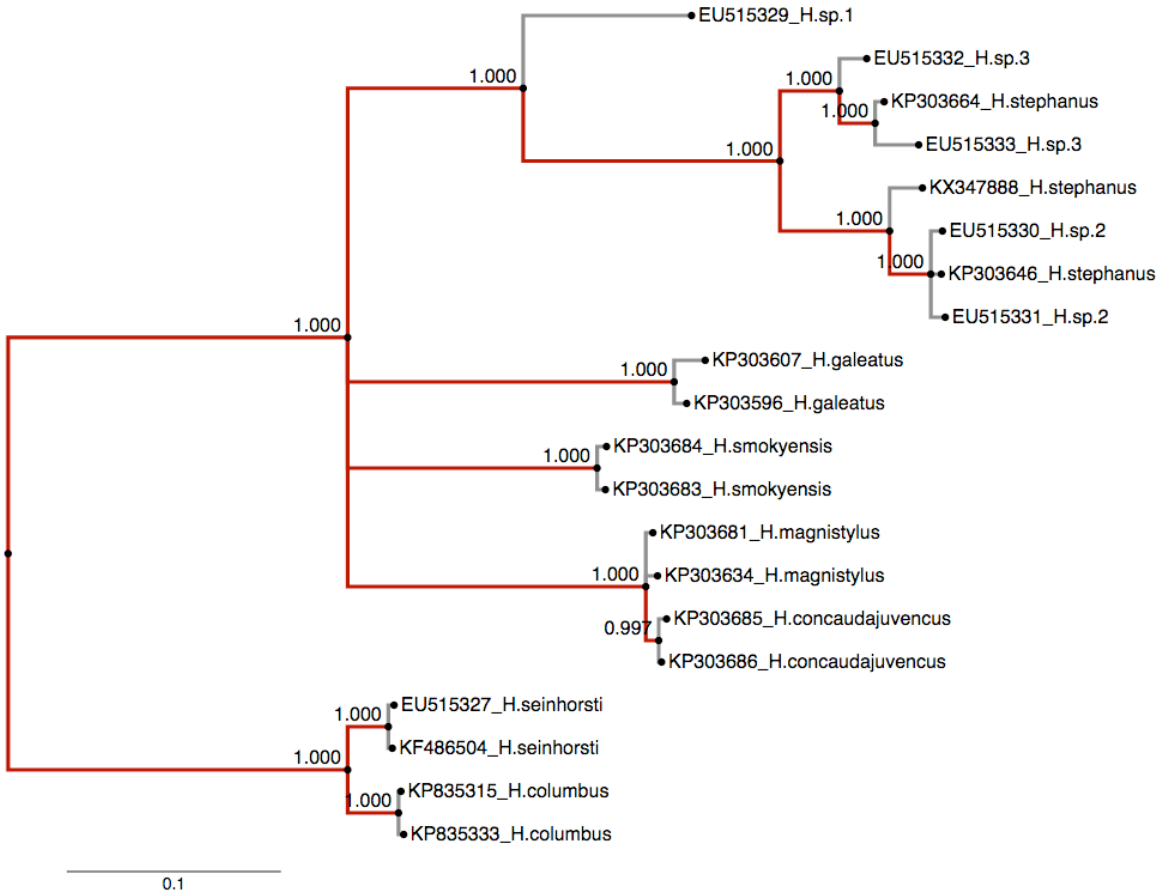

Fig. 1.6. Molecular phylogeny of *Hoplolaimus* species based on ITS1 DNA sequences. Bayesian Inference tree obtained with MrBayes. Unique sequences labeled. Model: GTR+I+G. MCMC = 10,000,000 generations.

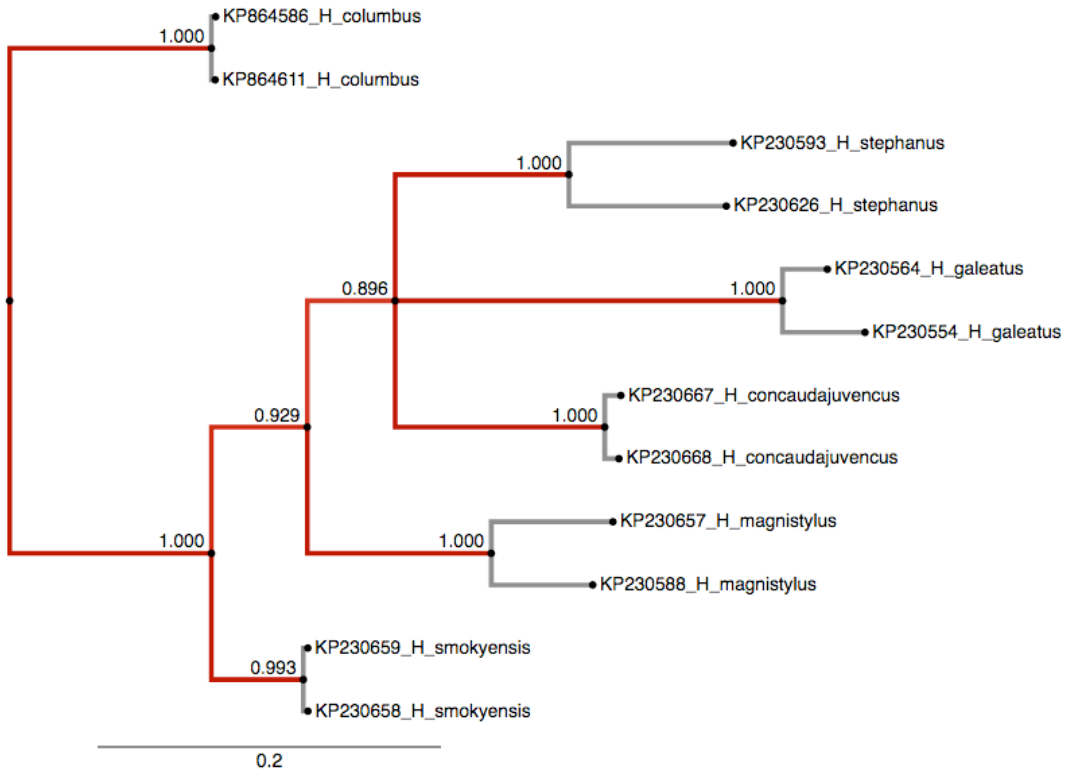

Fig. 1.7. Molecular phylogeny of *Hoplolaimus* species based on unique sequences of cytochrome oxidase subunit I (COI) DNA. Bayesian Inference tree obtained with MrBayes. Model: GTR+I+G. MCMC = 10,000,000 generations.

CHAPTER II.

ADDENDUM TO THE MORPHOLOGY OF *HOPLOLAIMUS COLUMBUS*

Introduction

Hoplolaimus columbus (Sher, 1963) was first collected by Q. L. Holdeman in 1957 from soybean (*Glycine hispida*) soil and cotton (*Gossypium hirsutum*) soil in South Carolina, and named the Columbia nematode. Sher elaborated morphological characters of *H. columbus* females in the original description, and gave prudent observation of esophageal gland nuclei as five distinct and a sixth obscure. The description of males remained unknown until 1973, when a population of *H. columbus* with a 1:60 ratio of male to female was found from a harvested soybean field near Holly Hill, South Carolina (Fassuliotis, 1974).

Esophageal glands of plant parasitic nematodes are noticeable not only due to secretion (Hussy et al., 2002), but also due to their significant roles in nematological taxonomy and evolutionary research (Handoo & Golden, 1992). In the order of Tylenchida, there are normally three esophageal gland cells. Each of two subventral gland cells contains one nucleus and so does the only dorsal gland cell. The subventral gland cells and the dorsal gland cell contain one nucleus each. It is unusual to find multiple nuclei in *H. columbus*, which has become an important characteristic for subgenus identification. However, the origination of multiple nuclei was under argument, either nuclei were duplicated from each cell (Fortuner, 1991), or one in the dorsal gland turned into four (Siddiqi, 2000).

Previous researches only provided drawings of *H. columbus* esophageal gland nuclei, as well as the male characteristics (Sher, 1963; Fassuliotis, 1974; Fortuner, 1991; Siddiqi, 2000). In this report, we present our observation in photographs of *H. columbus* esophageal gland cell nuclei, and utilize *H. galeatus* three nuclei for comparison. We also report one specimen of male *H. columbus* and show the first photograph of its characteristics under an optical microscope. Moreover, different from the usual round tail on adult females, we found two rare cases of tail malformation.

Materials and Methods

Hoplolaimus columbus were collected from a cotton field in Blackville, South Carolina. Nematodes were extracted from the soil by sugar centrifugal flotation (Jenkins, 1964).

Population species were identified by both morphological and molecular methods. DNA was extracted from individual nematodes handpicked from each population, using Sigma Extract-NAmp kit (XNAT2) (Sigma, St. Louis, MO). Species-specific primers were used in PCR for DNA amplification (Bae et al., 2008, Ma et al., 2011) to confirm the identification of *H. columbus* and *H. galeatus* populations.

Morphological observation was operated under an Olympus BX60 microscope, which was equipped with a camera and software iSolution Lite for photograph (Image and Microscope Technology i-Solution, Inc.). For esophageal gland nuclei observation, live nematodes were picked with a hair-needle and released in a drop of tap water on slides. Glass beads were added into the water to prevent nematodes from being crushed

by cover slides. In some slides, Cango red was added into the water for better contrast view.

The only male specimen of *H. columbus* was fixed in formalin-glycerol fixative (Golden in Hooper, 1970). Additional females from the same population were fixed as well. Observation and photography were operated as afore mentioned.

Results and discussion

Live nematodes in a drop of water on slides were observed directly for identifying esophageal gland nuclei. Both juvenile and adult *H. columbus* individual were investigated. Most *H. columbus* adults, if well developed, have a slim and opaque body. A viable environment is crucial for esophageal gland nuclei, since gland cells would quickly deform once nematodes begin to die. However, opaque bodies usually burden the observation of faint characteristics such as esophageal gland nuclei.

In Figure 2.1, A_1 shows six nuclei in gland cells of a juvenile *H. columbus*. Each arrow indicates one nucleus. The two posterior nuclei belong to the subventral glands and the other four nuclei are in the dorsal gland. The shape of the dorsal gland could be seen faintly as a long and narrow triangle. Sizes of all six nuclei are visually similar. For comparison, A_2 presents nuclei in *H. galeatus.* Three nuclei are located in two regions. One plain nuclei, larger than the other two, is located close to anterior end. The dorsal gland shape is similar to the one in A_1. Two smaller nuclei are near the intestine region and overlap.

B_1 and B_2 show the same adult *H. columbus*. The only difference between the two images is focus level. Cango red was added into water providing higher contrast. Esophageal gland cells have been squeezed tightly by the intestine. White arrows indicate four nuclei on the same focus level, while two black arrows indicate another two nuclei. B_2 only includes the four nuclei in focus. B_1 is taken slightly zoomed out from the B_2 level. B_1 clearly shows one clear nucleus overlapped upon the other four nuclei, as well as another faint shape of nuclei below the four nuclei.

In conclusion, these two photos present four nuclei in the same focus level and two other nuclei were separated by the four nuclei focus on each side. Although gland cells could not been differentiated, B_1 and B_2 support the observation of 4 dorsal gland nuclei in A_1, and indicate three different locations as those nuclei distributed.

Observation supported Siddiqi's statement that there are four dorsal esophageal gland nuclei of *H. columbus* and only one dorsal nucleus in *H. galeatus.* Sizes of nuclei observed in *H. columbus* were all similar, while in *H. galeatus* the dorsal nucleus was larger than subventral nuclei. In adults, the intestines usually push-on the esophageal gland cells anteriorly. This occurrence causes visual overlap of esophageal gland cells, which obstructs the observation of either gland cells or nuclei. This occurrence also supports Siddiqi's assumption that the "report of five nuclei in some species is an error since one of the two subventral gland nuclei is easily overlooked as the two are not in the same optical level," especially when the observation should be ideally performed with live and active nematodes.

Only one male, as Figure 2.2 was found from a population of 120 nematodes per cc soil in a cotton filed in Blackville, South Carolina. The male individual has 1370.5µm body length, 50.9 µm body width, 44.8 µm stylet length, and 43.8 µm spicule length. The entire body were diagnosed carefully and no lateral incisure was found. These measurements and characteristics are consistent with the original description (Fassuliotis, 1974), except the tail shape. The posterior end of the tail is blunter and shorter than of the original drawing. The bursa is smaller than the original description and beginning at a lower position of the anterior end of the spicules. The tail shape is also different from the sharp pointy shape of *H. galeatus* or *H. stephanus* males, which are more abundant. The females of *H. columbus* from the same population were observed and identified as well, and six of the females were fixed in formalin-glycerol fixative. Those females did not differ morphologically from the reported record.

Two cases of abnormal female tails in *H. columbus,* as in Fig 2.3, were observed from a cotton field in Blackville, South Carolina from 2017-2018. Female tails are folded inwards. This occurrence is similar as the one reported in the original description of *H. magnistylus* (Robbins, 1982), but it is the first report on *H. columbus*.

References

Bae, C.H., Szalanski, A.L., and Robbins, R.T. 2008. Molecular analysis of the lance nematode, *Hoplolaimus spp*., using the first internal transcribed spacer and the D1-D3 expansion segments of 28S ribosomal DNA. Journal of Nematology 40:201–209.

Fassuliotis, G. 1974. A description of males of *Hoplolaimus columbus.* Journal of Nematology 6:116–118.

Fortuner, R. 1991. The Hoplolaiminae. In: Nickle W.R. Manual of agricultural nematology. New York, USA, Marcel Dekker, Inc.

Handoo, Z.A. and Golden, A.M. 1992. A key and diagnostic compendium to the species of the genus *Hoplolaimus* Daday, 1905. Journal of Nematology 24:45–53.

Hooper, D.J. 1970. Handing, fixing, staining and mounting nematodes. In: Southey J.F. Laboratory methods for work with plant and soil nematodes. London, H.M.S.O. Pp:69–70.

Hussy, R.S., Davis, E.L. and Baum, T.J. 2002. Secrets in secretions: genes that control nematode parasitism of plants. Brazilian Journal of Plant Physiology 14:183–194.

Jenkins, W.R. 1964. A rapid centrifugal-flotation technique for separating nematodes form soil. Plant Disease Reporter 48:692.

Ma, X., Agudelo, P., Mueller, J.D. and Knap, H.T. 2011. Molecular characterization and phylogenetic analysis of *Hoplolaimus stephanus.* Journal of Nematology 43:25–34.

Robbins, R.T. 1982. Description of *Hoplolaimus magnistylus* n. sp. (Nematoda: Hoplolaimidae). Journal of nematology 14:500–506.

Sher, S.A. 1963. Revision of The Hoplolaiminae (Nematoda) II. *Hoplolaimus* Daday, 1905 and *Aorolaimus* N. Gen. Nematologica 9:267–296.

Siddiqi, M.R. 2000. Tylenchida parasites of plants and insects, 2nd Ed. Wallingford, UK, CABI Publishing.

Figures

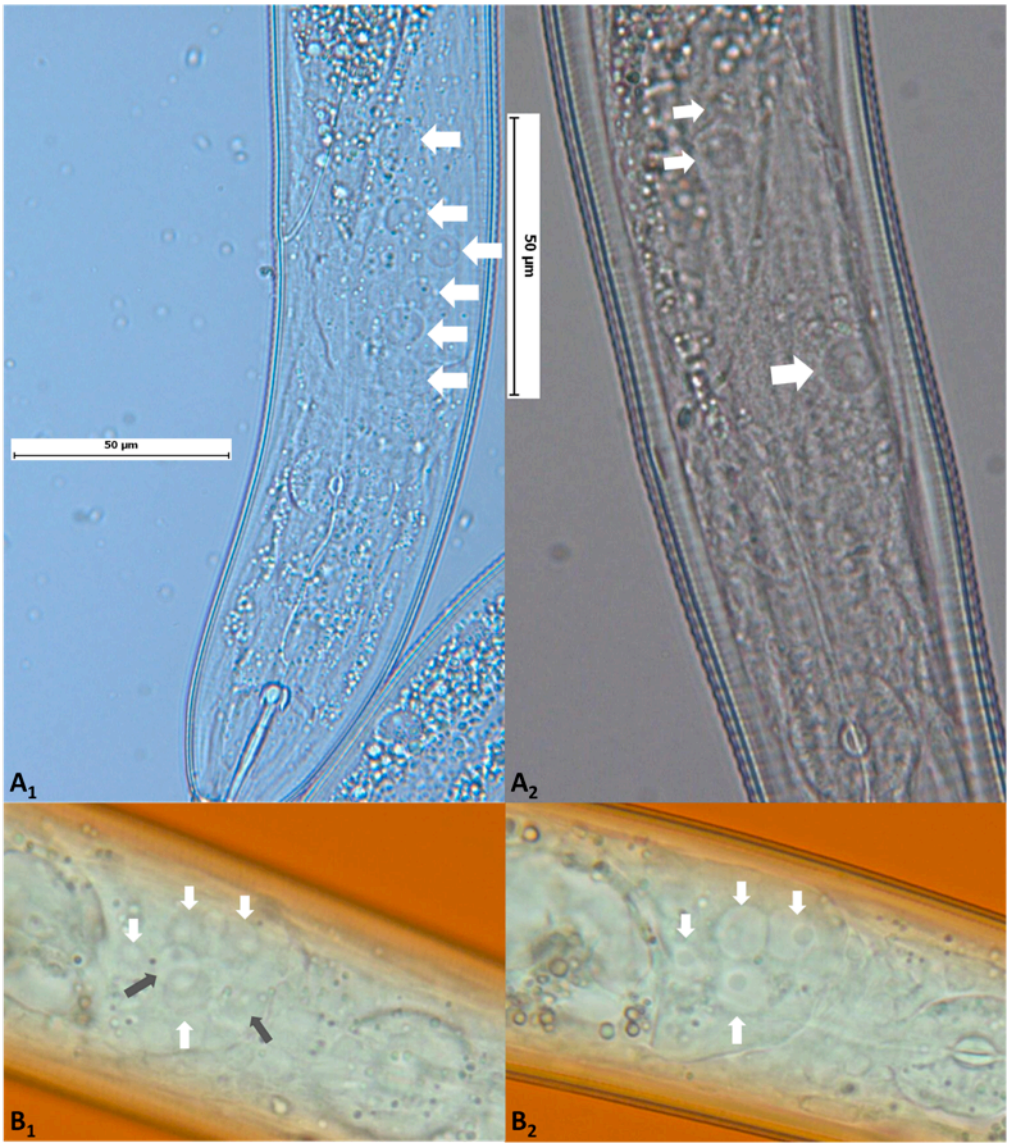

Figure 2.4. Esophageal gland nuclei observation. A_1 *H. columbus* with 6 nuclei. A_2 *H. galeatus*. B1 and B2: same *H. columbus* specimen with different focus. Arrows indicate esophageal gland nuclei.

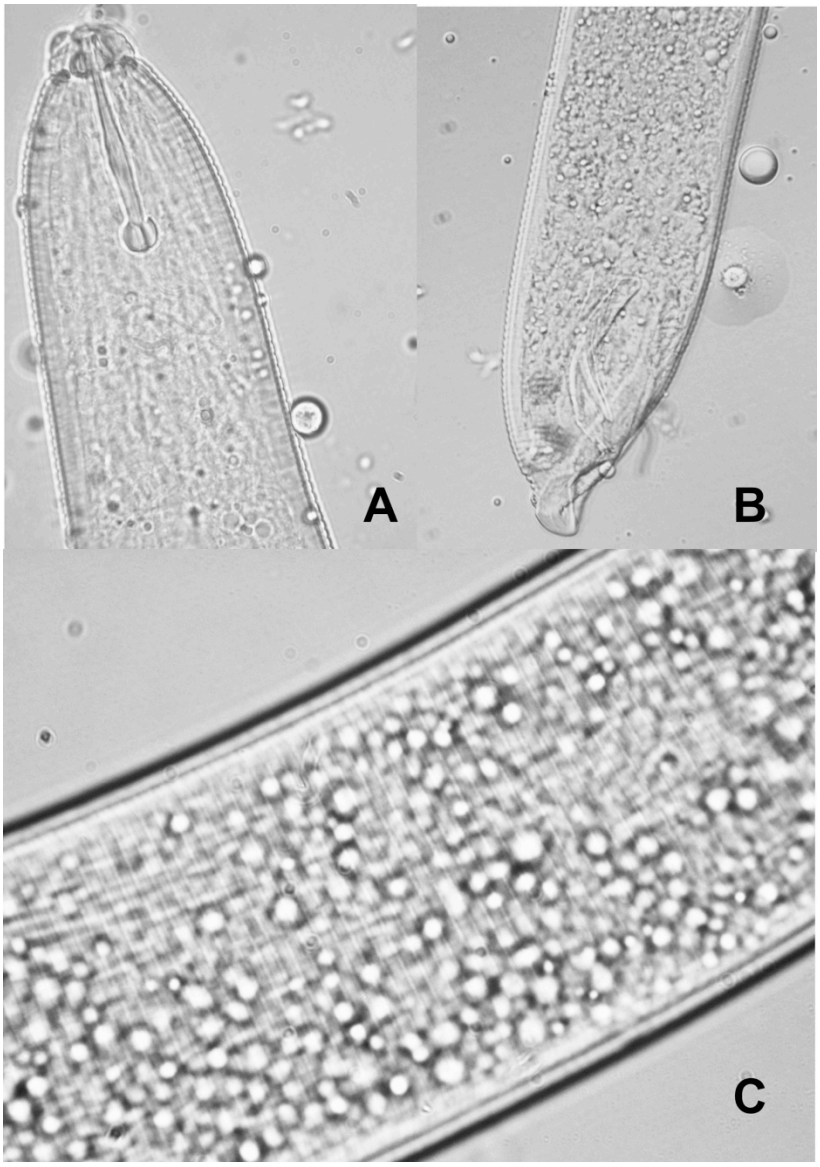

Figure 5.2. *H. columbus* male. A, Anterior region, B, tail region, C, body annulus without any lateral incisura

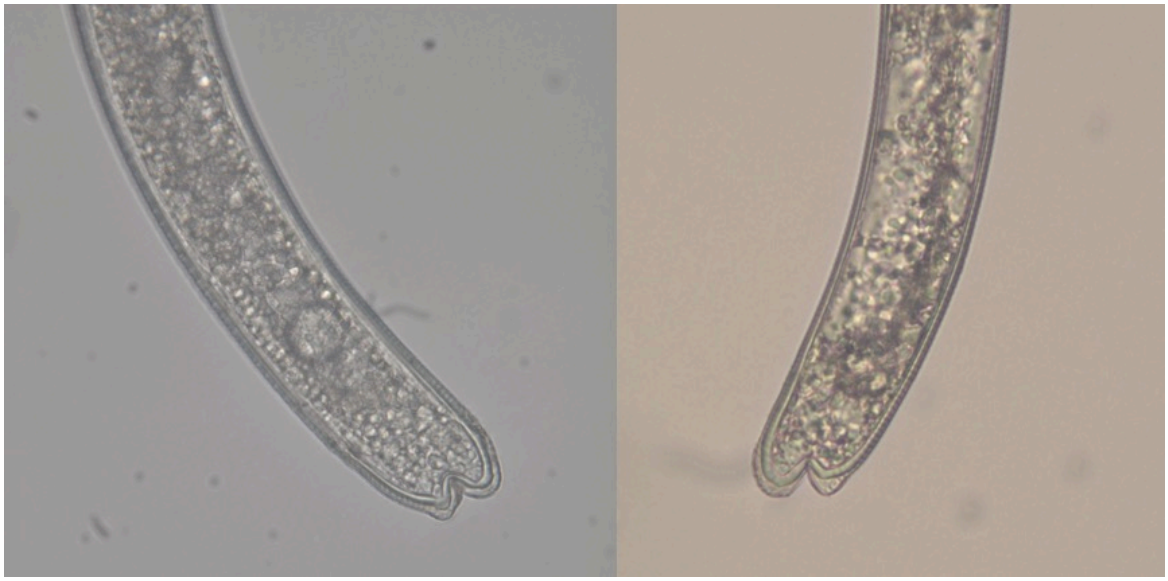

Figure 2.6. *H. columbus* female abnormal tails

CHAPTER III.

DE NOVO ASSEMBLY, COMPARATIVE ANNOTATION, AND PHYLOGENETIC ANALYSES REVEAL UNIQUE MITOGENOMIC CHARACTERISTICS OF *HOPLOLAIMUS COLUMBUS* IN NEMATODA PHYLUM

Data deposition: The mitochondrial genome sequences determined in this study were deposited into GenBank with the following accession numbers: *H. columbus* MH657221

Introduction

In the phylum Nematoda, plant-parasitic nematodes are distinguished from other free-living nematodes and animal parasitic nematodes by their mouthpart and stylet, which helps them to penetrate sturdy plant cell walls while digging and feeding. A number of plant-parasitic nematodes have been recognized as major pathogens that damage agricultural crops and artificial vegetation, which leads to annual losses of more than 150 billion dollars (Abad et al. 2008). In a survey performed in 1994 on a variety of crops in the US, the major genera of phytoparasitic nematodes reported to cause crop damage include 6 main genera (Koenning et al. 1999): cyst nematodes (*Heterodera* spp.), lance nematodes (*Hoplolaimus* spp.), root-knot nematodes (*Meloidogyne* spp.), lesion

nematodes (*Pratylenchus* spp.), reniform nematodes (*Rotylenchulus* spp.), and dagger nematodes (*Xiphinema* spp.). All other genera have mitochondrial genomes reported; however, *Hoplolaimus* spp. still have limited molecular references available.

Hoplolaimus spp. are migratory ecto-endo parasitic nematodes with a distinct cephalic region and a massive well-developed stylet (Sher 1963). They have a wide host range, including turf grasses, cereals, soybean, corn, cotton, sugar cane, and some trees. They usually feed on plant roots, migrate inside or surrounding the plant, destroy cortex cells, and result in the extension of necrotic lesions (Lewis and Fassuliotis 1982). *H. columbus*, also known as the Columbia lance nematode, is considered to be one of the most economically important species (Koenning et al. 1999). This species was first discovered in Columbia, South Carolina, and later was reported in North Carolina, Georgia, Alabama, and Louisiana (Astudillo and Birchfield 1980; Lewis and Fassuliotis 1982; Gazaway and Armstrong 1994). However, most of its morphologically close relatives are from Pakistan, India, China, and Japan (Siddiqi 2000). For molecular phylogenetic analysis, nuclear sequences, such as internally transcribed spacer sequences and 28S gene sequences, were utilized (Bae et al. 2008; Ma et al. 2011). The cox1 gene mitochondrial marker was used for genetic variability analysis (Holguin et al. 2016).

The mitochondrion is an organelle widely found in metazoans, which is thought to have descended approximately 1 billion years ago from the alpha-proteobacteria (Sagan 1967; Andersson et al. 2003; Boussau et al. 2004). It typically contains a single, circular molecule of DNA, while multiple circular mitochondrial molecules have been

reported (Armstrong et al. 2000; Suga et al. 2008; Philips et al. 2016). The mitochondrial genome (mt-genome) is relatively smaller than the nuclear genome of the organism and can also have a large range of genome sizes depending on different organisms. Most reported plant mt-genomes range between 200-2,000 kbp (Morley and Nielsen 2017), and for animals, the sizes are compact, typically approximately 16.5 kbp (Al Arab et al. 2017). The mt-genome size standard deviation of the most highly sampled metazoans (Chordata, Echinodermata, Arthropoda and Platyhelminthes) is relatively smaller than that of Mollusca, Porifera and the enoplean Nematoda (Gissi et al. 2008). Mt-genomes reported in the phylum Nematoda contain a large amount of gene content and genome size variation (Molnar et al. 2011). Each completed sequenced metazoan mt-genome normally contains genes encoding proteins, tRNAs and rRNAs, which include ATPase complex genes (*atp6* and/or *atp8*), cytochrome b gene (*cytb*), subunits I-III of cytochrome c oxidase (*cox1*, *cox2*, and *cox3*), and 7 subunits of the respiratory chain NADH dehydrogenase (*nad1, nad2, nad3, nad4, nad4l, nad5,* and *nad6*) (Chomyn and Attardi 1987). However, subsequent to the early reported mt-genomes of *Caenorhabditis elegans* and *Ascaris suum* (Okimoto et al. 1992), most of the genomes are missing *atp8*, except *Trichinella spiralis* (Lavrov and Brown 2001) and *Trichuris* spp. (Liu et al. 2012) in the order Trichinellida. Moreover, pseudogenes are present in *C. briggsae* and *Camallanus cotti* (Howe and Denver 2008; Zou et al. 2017), and repeated genes are present in the mitochondrial genomes of *Hexamermis agrotis*, *Romanomermis culicivorax* and *R. iyengari* (Beck-Azevedo and Hyman 1993; Lagisz et al. 2013).

In this study, we report the first complete mt-genome of the lance nematode *H. columbus*, sequenced by the whole-genome amplification and Illumina MiSeq method. Multiple assembly tools and parameter sets were tested for confirming the sequence reconstruction. Annotations were performed with both invertebrate mitochondrial genetic code 5 (Arthropoda, Mollusca and Nematoda) and code 14 (Platyhelminthes and Nematoda). In addition to the mt-genome of *H. columbus*, we investigated phylogenetic relationships among 92 unique nematode species using the maximum likelihood method. The alignment was created with concatenated nucleotide sequences of 12 mitochondrial protein-coding genes, except *atp8*, which is normally missing in the nematode mitochondrion. The gene order of all protein-coding genes was surveyed. Our study contributes to the evolutionary study of the Nematoda phylum and is also a tool for plant-parasitic nematode research. Of special significance are the potential contributions to the elucidation of the origin of *H. columbus*, first reported from South Carolina, USA, in 1963 but believed to be an introduced species to the United States.

Materials and Methods

Nematode samples and DNA extraction

Soil samples containing *H. columbus* specimens were collected from the Edisto Research Center in Blackville, South Carolina. Nematodes were extracted from soil samples by the sugar centrifugal flotation method (Jenkins 1964). Specimens were diagnosed by both morphological observation under an optical microscope following a

diagnostic key (Handoo and Golden 1992) and by molecular identification using the PCR method. Total DNA from each individual was extracted using a Sigma Extract-N-Amp kit (XNAT2) (Sigma, St. Louis, MO) and was probed using species-specific primers for species confirmation (Ma et al. 2011). The genome size of *H. columbus* was estimated by comparing with *C. elegans* cells by flow cytometry (Leroy et al. 2003). The results suggested that the whole genome of *H. columbus* is approximately 2-3 times that of the *C. elegans* genome and suggested that MiSeq was capable of covering an adequate depth for whole-genome sequencing.

Whole-genome amplification, library preparation and MiSeq sequencing

The live nematode specimens were merged in distilled water and starved for two weeks. Before DNA extraction, the nematodes were placed in a 3% hydrogen peroxide solution (Aaron Industry) for five minutes and then washed in distilled water three times to remove potential microorganism contamination from the nematode surface. Finally, the nematodes were placed in DNA Away solution (Molecular BioProducts, Inc.) to remove potential DNA and DNase contamination and were washed three times using PCR-grade water. Whole-genome amplification (WGA) was performed to obtain an adequate amount of total DNA of each individual nematode using an Illustra Ready-To-Go GenomiPhi V3 DNA Amplification kit (GE Healthcare). Three replications were performed, and the one with the best DNA concentration as tested by a Qubit fluorometer (Invitrogen) was selected for MiSeq library preparation. The library preparation followed the protocol of the Nextera XT kit (Illumina). The library concentration and fragment size

were respectively determined using a Qubit fluorometer (Invitrogen) and a Bioanalyzer 2100 (Agilent Technologies). The sequencing process used an Illumina MiSeq v3 kit, and more than 56 million 300 bp-reads were yielded. As a result, 98.1067% of the total reads were recognized as high-quality and passed the Illumina internal filter, and approximately 13 Gb of sequence data had a quality score (Q-score) greater than 30.

Mitochondrial genome assembly

Three methods were performed for the *de novo* assembly of the mitochondrial *H. columbus* genome. (1) NovoPlasty2.7.2 (Dierckxsens 2016): raw data first went through adapter trimming using Trimmomatic-0.36; the *de novo* assembly was then carried out by NOVOPlasty using the parameters of K-mer 31, insert range 1.6, and insert range strict 1.2. A partial cox1 gene sequence of *H. columbus* (KP864628) was used as a seed for initiating the assembly. (2) Geneious genome assembly: only 10% of the raw data was used for assembly. (3) MIRA (Chevreux et al. 1999) and MITObim (Hahn et al. 2013).

Mt-genome composition statistics, annotation and mt-genome map

The *de novo* assembled genome was studied using the Sequence Manipulation Suite (SMS) webserver (Stothard 2000) to obtain composition statistics. Tandem repeats were identified by Tandem Repeats Finer 4.09 (Benson 1999). Protein-coding genes and RNA genes of the confirmed mt-genome were predicted by the MITOS online webserver (Bernt et al. 2013), and RNAs were annotated using MITFI (Jühling, et al. 2012). Annotation results were checked using their quality values. The mt-genome map was

drawn by CGView Server (Grant and Stothard 2008). For GenBank submission and phylogenetic analyses, we chose mitochondrial genetic code 14, since this code was discovered from a plant-parasitic nematode species, *Radopholus similis* (Jacob et al. 2009), which is within the same taxonomic order as *H. columbus.*

Phylogeny reconstruction and protein-coding gene arrangement order

Phylogenetic relationships among the 92 species, as shown in the Table 3.4, were analyzed based on concatenated nucleotide sequences of 12 protein-coding genes (PCGs) (17340 bp alignment, all concatenated sequences following the order: atp6-cox1-2-3-cytb-nad1-2-3-4-4L-5-6). All PCG information of the 92 nematodes was downloaded from GenBank and reorganized using BioPerl (Stajich et al. 2002) and BioPython (Cock et al. 2009). The mt-genome of *Rotylenchulus reniformis* has no annotation information in GenBank; therefore, we annotated this genome using MITOS, with genetic code 14. The mt-genome of *Globodera pallida* has been reported (Armstrong et al. 2000), but it was not found in the GenBank database; therefore, it was not included in this study. The alignment of the 92 sequences was first performed with the MAFFT version 7 webserver (Kuraku 2013) and saved as a fasta file. The alignment files were then converted into a phylip file by the webserver Converter on Phylogeny.fr (Dereeper and Guignon et al. 2008; Dereeper et al. 2010). The maximum-likelihood phylogeny of the 92 species was performed on the PhyML3.0 webserver (Guindon et al. 2010). The number of substitution rate categories was set as 4, and the final model for ML analysis was GTR +G+I, which was selected by SMS (Smart Model Selection in PhyML) (Lefort et al.

2017). Nematode mt-genome sizes and PCG orders were surveyed from GenBank, collected and organized in Microsoft Excel.

Results

Assembly results and general features of the mt-genomes

Trimmed data were assembled by both NOVOPlasty2.5.9 and the updated version, NOVOPlasty2.7.2. A total of 40038720 reads were pooled, and 687794 reads were aligned. As a result, 110852 reads were finally assembled. The average coverage was 1476x. The genome was circularized automatically by the assembler. This assembly result was confirmed by two other methods: (1) the raw data were assembled by Mira4.02-MITOBim1.9; and (2) 10% of the raw data was randomly picked for assembly by Geneious11.1.2. The mt-genome of *H. columbus* is a closed-circular DNA molecule of 25228 bp in length, as Figure 3.1 which is the largest single chromosome in reported plant-parasitic nematodes. For *de novo* mitochondrial genome assembly, its nucleotide composition contained 4281 Gs (16.97%), 7178 As (28.46%), 11634 Ts (46.12%), and 2132 Cs (8.45%). The mt-genome composition is strongly biased towards A+T (74.57%). The GC skew and AT skew values were 0.3351 and -0.2369, respectively. All tandem repeats were found by Tandem Repeats Finder (Benson 1999). The consensus size of all tandems ranged from 6 bp (13.3 copies) to 237 bp (2.2 copies). Entropy values were calculated by the program for each tandem and ranged from 0.67 to 1.96 (Table 3.1).

Protein-coding gene (PCG) predictions

In our study, annotations were predicted using both invertebrate mitochondrial genetic code 5 (Arthropoda, Mollusca and Nematoda) and code 14 (Platyhelminthes and Nematoda) on the MITOS webserver, as in Table 3.2. Both annotations gave 12 PCGs and missed the atp8 gene after checking their quality values. The PCG codon usage from two genetic codes is listed in Table 3.3. The differences between the two codon usages are on AAA, ATA, and TAA. In genetic code 5, ATA codes Met, AAA codes Lys, and TAA is a stop codon. In genetic code 14, ATA codes Ile, AAA codes Asn, and TAA codes Tyr. The pattern of codon usage in the PCGs was biased towards T-rich codons, which has been described by Kim et al. (2017) as having two or more Ts per triplet. This pattern was seen in both annotation results. There are six codons having a usage number over 100, including 623 TTT (Phe, 17.43% of total usage), 279 TTA (Leu, 7.8%), 234 ATT (Ile, 6.54%), 220 TAT (Tyr, 6.15%), 124 TTG (Leu, 3.47%), and 103 GTT (Val 2.88%) (Table 3.3).

The PCG content of the two annotation results is listed in Table 3.2. Most gene sequences are discretely distributed throughout the mt-genome. Overlapped genes are bordered in the table. The two annotation results agreed that the cox1 gene was the longest (at 1413 bp), and the results also agreed on its locus. However, the start codon is AGC (Ser), and the stop codon is GCA (Ala), which are not in agreement with either genetic code. Both annotations agreed regarding two additional PCGs, the nad1 gene and the atp6 gene. These two genes use ATT as the start codon, which obeys genetic code 5,

but have not yet been reported in genetic code 14. Moreover, the ATG stop codon in the nad1 gene and GAA in the atp6 gene given by the results are in conflict with both genetic codes.

The differences between the two annotation results regarding PCGs are found in the cox2, nad3, nad5, nad2, cox3, nad4, cytb, nad6, and nad4l genes. For the cox2, nad3, cox3 and cytb genes, both annotations agree on the stop loci but disagree on the start codons. The stop codons used by these genes are CTA, TGG, TGA and ATT; however, none of them obey either of the two genetic codes. The start codons of ATT, ATA, and TTG, used by these genes in the code-5 results, obey code 5. However, in the code-14 results, only the start codon ATG in the cytb gene obeys code 14. Other start codons of GTT, TTT, and TAT have not yet been reported in code 14. For the nad5 gene, both annotation results agree on the start/stop codons as ATA/AAT; however, the gene loci and lengths are different between them (Table 3.2). The nad2 and nad4 genes in the code-5 results, using ATT and TTG as start codons, obey code 5 but use TTT and TGA as stop codons, which is against the code. In the code-14 results, both the nad2 and nad4 genes use TTA as a start codon, which has only been reported for genetic code 4, and AAA and TTT as stop codons have not yet been reported in this code.

The nad6 and nad4L genes are different in both loci and codons. In the code-5-result, the nad6 gene has start/stop codons of ATT/TTA and is located on the plus strand, while in the code-14-result, the nad6 gene has start/stop codons of GAT/GTT and is located on the minus strand. For the nad4L gene, the code-5 result shows the start/stop

codons as ATT/GTC, and it is next to the nad6 gene before a long noncoding region, while it is located after the noncoding region, with start/stop codons of ATT/ATG and close to the nad1 gene in the code-14-result. Disagreement in the start/stop codons is also found, as previously described. The alignment of each gene from the two annotation results is presented in Figure 3.2.

Ribosomal RNA and transfer RNA gene predictions

Two ribosomal RNA genes were predicted by each genetic code. The gene loci and sizes agreed in both codes (Table 3.2), with an rrnS length of 70 bp and an rrnL length of 529 bp. These two rRNAs are much smaller than in other plant-parasitic nematodes (674 bp and 817 bp in *Globodera ellingtonae*, 673 bp and 806 bp in *Heterodera glycines*, 610 bp and 804 bp in *Meloidogyne arenaria*, 601 bp and 789 bp in *Meloidogyne chitwoodi*, 610 bp and 804 bp in *Meloidogyne incognita*, 692 bp and 840 bp in *Radopholus similis*, and 686 bp and 895 bp in *Pratylenchus vulnus*) (Jacob et al. 2009; Gibson et al. 2011; Sultana et al. 2013; Humphreys-Pereira and Elling 2014; Phillips et al. 2016). Although the rrnS and rrnL are both abnormally smaller, their BLAST results in GenBank respectively match the rrnS and rrnL genes of other nematodes in the Tylenchida order, and even nematodes from other orders, such as *Oesophagostomum dentatum* (700 bp and 959 bp), *Bursaphelenchus xylophilus* (700 bp and 948 bp), *Cylicocyclus nassatus* (698 bp and 974 bp), and *Setaria digitate* (680 bp and 967 bp). The predicted structures are shown in Fig 3.3 and 3.4.

Both results present 20 tRNAs (from 50 bp to 73 bp) but differ in their partial contents. Two results agree on 16 tRNAs (C, D, E, F, G, H, L1, L2, N, P, Q, R, S1, S2, V, W) and are all missing trnA and trnT. With genetic code 5, the result contains trnI and a duplicated trnW gene but is missing trnM. With genetic code 14, the result has a trnM and a duplicated trnN gene, while the trnI gene is missing. Moreover, the gene loci of trnK and trnY did not agree between the two annotation results.

Structure predictions of tRNAs from the two annotation results are presented in Figures 3.3 and 3.4. Most reported nematode tRNAs are not like human mitochondrial tRNAs, which have a canonical cloverleaf structure. Some of the nematode tRNAs lack the T-arm, and some are missing both arms (Jühling et al. 2012; Palomares-Rius et al. 2017). In *H. columbus*, all predicted tRNA structures have at least one arm; however, most of them have abnormal structures predicted. For both annotation results, the variable loop was found on the acceptor stem (trnC and E), on the T-stem (trnC, R, S1, and V), and on the anticodon arm (trnE, F, G, R, V, and Y). The T-arm is missing in trnE, G, H, L1, L2, P, V, and Y. The D-arm is missing in S1, S2, and N (aat). The structure of trnW (tga) has a T-stem but no T-loop. Moreover, in the code-5 result, trnW (tgg) is missing the T-stem, and in the code-14 result, trnN is missing the entire T-arm. The predicted structures of trnK (aaa) (code-5 result) and trnN (aaa) (code-14 result) have the same structure of a canonical cloverleaf and are from the same sequence locus but differ in their genetic code usage.

Noncoding region prediction

In the code-5 result, the total noncoding region length is 14955 bp. In the code-14 result, the total noncoding region length is 15115 bp. Both annotation results indicate a large noncoding region (over 7000 bp) ending with a trnR gene. In the code-5 result, the large noncoding region has 7685 nucleotides and contains 1494 Gs, 2291 As, 3064 Ts, and 835 Cs. The respective percentage of each is 19.44%, 29.81%, 39.87%, and 10.87%. The AT content is 69.68%. In the code-14 result, 7307 nucleotides of the large noncoding region contain 1433 Gs, 2186 As, 2897 Ts, and 790 Cs. The respective percentage of each is 19.61%, 29.92%, 39.65%, and 10.81%. The AT content is 69.57%. Although the noncoding region in the code-14 result is 5.08% smaller than that of the code-5 result, the AT content remains relatively constant and is only 0.16% smaller.

In this large noncoding region, multiple tandem repeats were found (from 17188 to 21165). The longest tandem period has 238 bp, with a copy number of 2.2, a consensus size of 233 bp, and an entropy of 1.56 (Table 3.1). Two additional long tandems were found to be over 100 bp. In this noncoding region, all entropies have higher values (1.49-1.96) than the rest of the genome (0.66-1.29), except for one region (11990-12210, entropy 1.91). Most tandems in this region have a strong bias towards Ts (46%-62%) and only a few Cs (1%-6%), while the tandem with the highest entropy has only 25% T but the highest number of Cs (18%).

H. columbus has the largest total noncoding region in a single chromosome when compared with other reported Tylenchida nematodes (*Pratylenchus vulnus* 8821 bp,

Radopholus similis 3984 bp, *Globodera ellingtonae* chromosome I 9147 bp and chromosome II 10347 bp, *Meloidogyne chitwoodi* 5896 bp, and *Meloidogyne incognita* 5311 bp, 80.8%) (Phillips et al. 2016). When comparing the longest noncoding regions, *H. columbus* has a unique length and AT content compared with other long noncoding regions and AT contents, such as those of *Pratylenchus vulnus* (6858 bp, 73.0%), *Radopholus similis* (3705 bp, 87.3%), *Globodera ellingtonae* chromosome I (8059 bp, 62.0%) and chromosome II (7232 bp, 62.3%), *Meloidogyne chitwoodi* (5404 bp, 86.1%), and *Meloidogyne incognita* (4097 bp, 80.8%) (Phillips et al. 2016).

Phylogenetic analyses

We conducted maximum likelihood phylogenetic analysis of 17340 nucleotide characters from concatenated sequences of 12 protein-coding genes of 92 nematode species, including the newly sequenced *H. columbus.* The outgroup was set by 4 species of nematodes in which the atp8 gene was found in their mitochondrion (Lavrov and Brown 2001; Liu et al. 2013), though sequences of the atp8 genes were not included in the phylogenetic analysis. The phylogenetic tree is presented in Figure 3.5. Phylogenetic relationships were mainly consistent with previous studies of mitochondrial genomes, with strong bootstrap percentages (Park et al. 2011; Sultana et al. 2013; Humphreys-Pereira and Elling 2014; Phillips et al. 2016; Palomares-Rius et al. 2017). All nematodes from the class Enoplea were grouped into one of three clades: mammal-parasitic (Trichinellida), plant-parasitic (Dorylaimida), or insect-parasitic (Mermithida), with strong nodal support (0.99-1.0). The class Chromadorea, including suborders Tylenchina,

Spirurina, Rhabditina, and Strongylida, is mainly grouped by the taxonomic orders or suborders.

H. columbus, as a migratory ecto-endoparasitic species, is a sister taxa of the monophyletic group that includes another three sedentary parasitic species: *Heterodera glycines*, *Rotylenchulus renifomis* and *Globodera ellingtonae*. All five *Meloidogyne* spp. are grouped in the same clade, to which *Radopholus similis* is a sister taxa. *Pratylenchus volnus* is at a distant position from all other taxa in the monophyly of the Tylenchida order, which is slightly different from its phylogenetic locus in several other reports (Sultana et al. 2013; Humphreys-Pereira and Elling 2014). All plant-parasitic nematodes of the Chromadorea class are mainly in the same monophyletic group, except for *Aphelenchoides besseyi* and *Bursaphelenchus xylophilus.* However, Siddiqi (1980) proposed, diagnosed and defined the order Aphelenchida and traced its origin and phylogeny; these two plant-parasitic nematodes belong to the Aphelenchida order. Moreover, these two species behave differently from other plant-parasitic nematodes. *Aphelenchoides besseyi* can feed on fungi in addition to plants, and *Bursaphelenchus xylophilus* has a life cycle that involves insect host participation. Our results support Siddiqi's conclusion of an independent order of Aphelenchida. The monophyly of Aphelenchida has also been presented in other published reports (Park et al. 2011; Sultana et al. 2013; Kim et al. 2015; Kim et al. 2017) but was discussed as part of the Tylenchomorpha infraorder. It may be ambiguous as to why species from the same infraorder were distantly separated in phylogenetic analyses.

Protein-coding gene order arrangements and genome size comparison

Although we found discrepancies between the two annotation results, the protein-coding gene order arrangements were identical. In Figure 3.6, we list all PCG arrangements of the 92 species and their genome sizes in alignment with their phylogenetic relationships. All PCG arrangements were aligned starting from the cox1 gene. In the class Enoplea, all Trichinellida nematodes have the atp8 gene, and all species have the same PCG order. *Trichinella spiralis* has a relatively larger mt-genome size, slightly above 16 kbp, and the genome sizes of the other three species are less than 16 kbp. In Mermithida, the PCG order is species-specific. *Hexamermis agrotis* has three continuous copies of the atp6 gene. *Romanomermis iyengari* and *Romanomermis culicivorax* have multiple copies of the nad3 gene.

Nematodes in Dorylaimida have a relatively smaller mt-genome size compared to the other species. All five species have a mt-genome size smaller than 14 kbp. *Paralongidorus litoralis* and *Longidorus vineacola* share two identical PCG order regions (cytb-nad4l-nad3-cox1 and cox2-cox3-nad2). The differences in these two PCG orders likely resulted from the partial reversed sequences (from atp6-nad4 to nad4-atp6 and from nad5-nad6-nad1 to nad1-nad6-nad5). *Xiphinema rivesi* and *X. americanum* have an identical PCG order but only share partial PCG order with *X. pachtaicum* (nad5-nad6, cytb-atp6-nad4, and cox3-nad1-cox1). The other two regions (md4l-nad3 and cox2-nad2) of *X. pachtaicum* were both reversed from the order of the two other *Xiphinema* spp.

In Tylenchida, nematodes generally have relatively larger genome sizes than those in the Chromadorea class. *Globodera ellingtonae* has two chromosomes, and the total size of the mt-genome is 32122 bp. *H. columbus* has the largest chromosome at 25228 bp. *Rotylenchulus reniformis* has a size of 24572 bp. *Heterodera glycines* has only a partial genome available, but the size is already close to 15 kbp, and it may contain large noncoding regions as well (Gibson et al. 2011). *Radopholus similis* has 16791 bp, which is the smallest complete mt-genome so far in Tylenchida but is slightly higher than normal animal mt-genomes of 16.5 kbp (Al Arab et al. 2017). *Meloidogyne* spp. have mt-genome sizes ranging from 17 kbp to 19.6 kbp, also higher than the normal size. PCG orders in Tylenchida are phylogenetically informative. The PCG orders of all five *Meloidogyne* spp. are identical. *Heterodera glycines* and *Rotylenchulus reniformis* share the same PCG order. Moreover, in all Tylenchida species, "atp6-nad5" suggests a conservative region. This region is unique to the Tylenchida order and has not been found in any other order of nematodes.

Wellcomia siamensis, *Enterobius vermicularis*, and *Passalurus ambiguous* are 3 vertebrate animal-parasitic nematodes infecting humans, pigs, and rabbits (Zhang et al. 2015), and they share the same PCG order. From *Heliconema longissimum to Brugia malayi* on the phylogenetic tree, there are 12 species having the same PCG order. In total, 15 species are monophylic, and 3 conservative regions of PCG order could be found (nad1-atp6, nad6-cytb-cox3, and nad3-nad5). All 15 species have a smaller genome size than Tylenchida (13590 bp – 14128 bp).

From *Philometroides sanguineus* to *Cylicocyclus nassatus* at the posterior end of the tree, most species follow the same PCG order as *Caenorhabditis elegans* and *Ascaris suum* (Okimoto et al. 1992). The unusual PCG orders are presented by two groups. One group is the *Strongyloides* genus, which has species-specific PCG orders for each. Phylogenetic relationships indicate that *S. ratti* and *S. stercoralis* are two close species, while *S. papillosus* is close to *S. venezuelensis*. Conservative regions of PCG order support these phylogenetic results. In *S. ratti* (16609 bp) and *S. stercoralis* (13751 bp), nad4-cox3-atp6-nad2 and nad3-nad1 were found to be two conserved regions. On the other side, those two conserved regions could not be found in *S. venezuelensis* (15956 bp) or *S. papillosus* (13909 bp), but the nad5-nad4 region was found to be conserved.

The other group has *Philometroides sanguineus, Gnathostoma spinigerum, Rhigonema thysanophora,* and *Heterorhabditis bacteriophora.* In the PCG order of *Philometroides sanguineus* (14378 bp), the region of nad1-atp6 exchanged its locus for nad2-cytb-cox3-nad4. In the PCG order of *Heterohabditis bacteriophora* (18128 bp), the region of cytb-cox3-nad4 was inserted in between the positions of nad3 and nad5. In the PCG orders of *Gnathostoma spinigerum* (14079 bp) and *Rhigonema thysanophora* (15015 bp), 2 regions are conserved (nad6-nad4l-cox1-cox2-nad3 and nad2-cytb-cox3) and 4 genes were shuffled (nad5, nad4, nad1, and atp6).

The survey results of the PCG orders and genome sizes were generally consistent with the phylogenetic relationships between the 92 nematode species. The multiple PCG

copies, order conservation, and order rearrangement is interesting and may help elucidate the evolution of these nematodes. The most reported genome sizes for nematodes are typically less than 16.5 kbp and range from Dorylaimida (12489-13519 bp) to Mermithida (15546-26194 bp) and Tylenchida (16791-25228 or 32122 bp). These results raised the hypothesis that in nematode taxonomic orders, the genome size is related to the PCG diversity, in other words, a larger mt-genome size is related to a higher PCG diversity in the Nematoda phylum. However, to answer this question, more mt-genome information and careful statistical methods must be included in the analysis.

Discussion

Nematode genome sequencing

In this study, we *de novo* sequenced and assembled the mitochondrial genome of *H. columbus* using whole-genome amplification (WGA), Illumina MiSeq, and bioinformatic methods. Two crucial aspects influence the success of the sequencing. The first is to obtain enough DNA from the nematodes. Mitochondrial DNA was extracted as part of the whole-genome DNA from the nematodes, and then long-PCR or step-out PCR was performed to amplify the mtDNA only (Gibson et al. 2011; Sultana et al. 2013) to obtain an adequate amount of DNA for sequencing. However, this method relies on

species-specific mitochondrial primers. Another way of obtaining enough DNA material is to extract the mitochondria from a large number of specimens (Jacob et al. 2009). However, to ensure a genetically homogeneous source of the DNA, this method requires a large number of nematodes from either a pure culture or a clonal population to reach the required amount, which is a high-cost method. The WGA method has been used for microsatellite genotyping and population structure determination nematode research (Arias et al. 2011; Kikuchi et al. 2016). It has been proven to be an efficient and effective method. However, one aspect is that WGA mixes fragments of nuclear DNA and mitochondrial DNA. To avoid potential issues of nuclear copies of mitochondrial DNA (numts), a careful bioinformatic protocol is required. In our study, three mt-genome assemblers were tested and showed consistent results, with depths of over 1000x. The genome size should also be considered before sequencing. Since different facilities and kits have different sequencing abilities and costs, genome size estimation was performed before deciding to use MiSeq in our study. WGA techniques have been widely used in single-cell sequencing for helminths and nematodes (Young et al. 2012; Holroyd and Sanchez-Flores 2012; Gawad et al. 2016) to solve the issue of DNA material deficiency. However, additional steps fragment the DNA material before the whole-genome sequencing step, and the sequencing results are inevitably poor. This fragmentation, therefore, leads to a heavy reliance on bioinformatic methods; in our study, we tried multiple methods of different computing principles to improve the reliability of the *H. columbus* mt-genome. Another limitation of our method is in defining the number of chromosomes. In our study, only one mitochondrial chromosome was assembled and

circularized. However, there are reports of multipartite mt-genome or multiple chromosomes in *Globodera* spp. (Armstrong et al. 2000; Phillips et al. 2016), which was monophyletic with *H. columbus*. Other multipartite mt-genomes have been recently reported, including *Rhabditophanes* sp. and *Ruizia karukerae* (Hunt et al. 2016; Kim et al. 2018). Considering the genomic diversity in the Tylenchida order, we hope to examine future research methods for generating additional plant-parasitic nematode genomic information.

Mitochondrial annotation

The number of nematode genomes that have been sequenced is rapidly increasing (Kumar et al. 2012), and some annotation pipelines have become available. Some characteristics of a reliable annotation pipeline are that it is fast, automatic, accurate and reproducible (Al Arab et al. 2017). There are several preliminary tools for annotating mt-genomes. DOGMA (Wyman et al. 2004) automates the annotation of extranuclear organelle genomes and has been used in many published nematode mt-genomes (Kim et al. 2006, Webb and Rosenthal 2011; Kim et al. 2018); however, it is not under active development, there will not be further updates, and it is unsupported, according to a recent announcement on the web. MITOS (Bernt et al. 2013) is designed to compute consistent *de novo* annotations of metazoan mt-genomes; MitoFish and Mitoannotator are designed specifically for aquatic animals (Iwasaki et al. 2013); Mitofy is developed for plant mitochondrial sequences (Alverson et al. 2010); and GeSeq can annotate mitochondria genomes but requires the manual setting of mt-genome references (Tillich

et al. 2017). Moreover, the use of unusual genetic codes (Hyouta et al. 1987), noncanonical codons (Nagaike et al. 2005; Jacob et al. 2009) and the existence of overlap between genes (Wolstenholme 1992) affects the annotation results significantly. In the published annotations of plant-parasitic nematodes, most of them were annotated using genetic code 5 (Sultana et al. 2013; Humphreys-Pereira and Elling 2014), though genetic code 14 was reported earlier in the research of *Radopholus similis* (Jacob et al. 2009). Some publications do not mention or discuss which genetic code they used, but they are considered to have used code 5 after we verified their reports (Gibson et al. 2011; Phillips et al. 2016). Most of the recently published studies did not discuss or mention the reasons for choosing genetic code 5 for annotating the mt-genomes, while they did notice that some of their annotation results (start codon or stop codon) did not fully obey the genetic code (Kim et al. 2016; Park et al. 2011; Sultana et al. 2013; Humphreys-Pereira and Elling 2014; Phillips et al. 2016; Palomares-Rius et al. 2017).

An additional issue of the accuracy of the annotation pipeline is the causality between annotation results and genetic codes and codons. For most published mt-genomes, the annotation is a one-way result referred from other genomes. The taxonomy and phylogeny are the fundamental knowledge supporting the referral decision. In this study, we considered the phylogenetic relationships between *H. columbus* and all other nematodes and tried multiple genetic codes for annotation of the *de novo* assembly using not only codes 5 and 14 but also code 9 (the echinoderm and flatworm mitochondrial code) and code 13 (the ascidian mitochondrial code). The results from code 9 are

identical to those from code 14, and the results from code 13 are identical to those from code 5. Once uploaded to GenBank, these published annotation results, including our sequences, will become references for other annotations, and the potential for bias will be increased. One way to experimentally increase the annotation accuracy is to extract mitochondrial RNAs (Jacob et al. 2009). RNA-seq might be another way to improve the accuracy, depending on the precise molecular operations, but it might be difficult to use for nematodes that are difficult to culture. Since annotation upon references is still a prevalent method, the importance of phylogeny is increased, and additional mt-genomes and RNA information from diverse sources are needed.

Nematode mitochondrial phylogeny

Prior to molecular phylogeny, morphological characteristics were the key evidence for forming hypotheses of phylogenetic relationships of nematodes. After nuclear gene sequences were implemented, such as 18S, 28S, and ITS DNA sequences, the classification system of nematodes began to be modified, and a new phylogenetic system was suggested (Blaxter et al. 1998; De Ley and Blaxter 2002; Holterman et al. 2006; van Megen et al. 2009). Meanwhile, mitochondrial DNA has been frequently used for phylogenetic and evolutionary studies because of their peculiar mutation characteristics and dynamics (Boore 2006), and the phylogeny reconstructed by mtDNA could disagree with that from nuclear genes (Shaw 2002). Moore (1995) stated that mtDNA may be more conservative than nuclear genes, based on topological analysis. He found that mtDNA could be more robust in describing the species than nuclear genes

when a similar amount of data was available. However, Springer et al. considered nuclear genes to be more efficient in reconstructing deep-level phylogenetic relationships (Springer et al. 2001). For prior results of nematode phylogeny, Park et al. (2011) found discrepancies between mtDNA trees and SSU rDNA trees (Nadler et al. 2007). In our study, the mitochondrial phylogenetic relationships of the 92 nematodes agree with most published results of the mitochondrial phylogeny of nematodes (Park et al. 2011; Sultana et al 2013; Kim et al. 2017). *H. columbus* is monophylic with other Tylenchida nematodes. In this research, we also included several additional species of Dorylaimida, which are also plant-parasitic nematodes. All Dorylaimida spp. are grouped in an independent clade, which solved the previously mentioned issue that the unstable position of *X. americanum* was possibly due to a single long-branch attraction (Park et al. 2011; Kim et al. 2015).

Amino acid sequences have been frequently used to study nematode phylogeny (Park et al. 2011, Kim et al. 2017). However, this method is rooted in annotation results, and therefore, potential bias from previous annotations can accumulate and influence the results. In our study, we addressed the issue of using different genetic codes, which limits the validity of using the amino acid sequence. In addition to nucleotide or amino acid sequences, gene orders are also considered phylogenetically informative (Sankoff et al. 1992). There are also methods to draw phylogenetic relationships based merely on gene orders (Bourque et al. 2002; Hu et al. 2014; Zhou et al. 2017). However, such phylogeny cannot use bootstrapping to assign confidence values, since the gene order is considered

as one character with multiple states (Moret and Warnow 2005; Shi et al. 2010). Further studies of phylogenetic analyses of mt-genome order would be interesting. In this study, we filtered RNA genes out of the gene order survey to ensure the accuracy of the information used from the annotation results. However, less could be more. Since tRNA genes have a significantly smaller size than PCG genes, the small sequences could become phylogenetic noise, thereby interrupting the visual inference from the arrangements. Whether to include RNAs or not is another issue to consider when using the gene order phylogeny.

In our study, the nematodes presented a high diversity of gene order among the different taxonomic orders. Furthermore, within the same order, nematodes present relative conservativeness in some PCG order regions. The phenomenon of nematode mt-genome order rearrangement is also unique when comparing with other organisms such as fungi (Aguileta et al. 2014), insects (Cameron 2014), plants (Tanifuji et al. 2016), and birds (Singh et al. 2008).

Conclusion

In this study, we *de novo* assembled the mitochondrial genome of *H. columbus*. This genome data fills an important blank in the knowledge of agricultural pathogens, adds a unique plant-parasitic nematode to the nematode phylogeny, and contributes to *Hoplolaimus* spp. research on population genetics, nematode speciation, and

biogeographic inference. We also elucidated the unique characteristics of the *H. columbus* mt-genome and addressed issues related to the annotation method used and nematode phylogeny. We hope that these data will support further advances in understanding the origins of nematode parasitism, their genomic dynamics, and their biological adaptations.

References:

Abadm P, et al. 2008. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nature Biotechnology. 26:909-915.

Aguileta G, et al. 2014. High variability of mitochondrial gene order among fungi. Genome Biol Evol. 6: 451-465.

Al Arab M, et al. 2017. Accurate annotation of protein-coding genes in mitochondrial genomes. Mol Phylogenet Evol. 106:209-216.

Alverson et al. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27, 1436-1448.

Andersson SGE, Karlberg O, Canbäck B, Kurland CG. 2003. On the origin of mitochondria: a genomics perspective. Philos Trans R Soc Lond B Biol Sci. 358: 165–179.

Arias RS, Stetina SR, Scheffler BE. 2011. Comparison of whole-genome amplifications for microsatellite genotyping of *Rotylenchulus reniformis*. Electronic Journal of Biotechnology. 14:3.

Armstrong MR, Blok VC, Phillips MS. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. Genetics. 154:181-192.

Astudillo GE, Birchfield W. 1980. Pathology of *Hoplolaimus columbus* on sugarcane. Phytopathology. 70:565.

Bae CH, Robbins RT, Szalanski AL. 2009. Molecular identification of some *Hoplolaimus* species from the USA based on duplex PCR, multiplex PCR and PCR-RFLP analysis. Nematology. 11:471–480.

Beck-Azevedo JL, Hyman BC. 1993. Molecular Characterization of Lengthy Mitochondrial DNA Duplications from the Parasitic Nematode *Romanomermis Culicivorax*. Genetics. 133: 933–942.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27: 573-580.

Bernt A, et al. 2013. MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation. Molecular Phylogenetics and Evolution. 69:313-319.

Blaxter ML, et al. 1998. A molecular evolutionary framework for the phylum Nematoda. Nature. 392: 71-75.

Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol. 21:439-446.

Bourque G, Pevzner PA. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res. 12: 26-36.

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG. 2004. Computational inference of scenarios for alpha-proteobacterial genome evolution. Proc Natl Acad Sci USA. 101:9722-9727.

Cameron SL. 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu Rev Entomol. 59:95-117.

Chomyn A, G. Attardi G. 1987. Mitochondrial gene products. Curr Topic Bioenergetics. 15:295-329.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 99: 45-56.

Cock PA, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25:1422-1423.

De Ley P, Blaxter ML. 2002. Systematic position and phylogeny. In: The Biology of Nematodes, D.L. Lee, ed., London: Taylor and Francis. p.1–30.

Dereeper A, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:W465-469.

Dereeper A, Audic S, Claverie JM, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol. 10:8.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.

Fortuner R. 1991. The Hoplolaiminae. In: Nickle WR, editor. Manual of Agricultural Nematology. New York: Marcel Dekker, Inc. p. 669–719.

Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 17:175-88.

Gazaway WS, Armstrong. 1994. First report of Columbia lance nematode (*Hoplolaimus columbus*) on Cotton in Alabama. Plant Dis. 78:640.

Gibson T, et al. 2011. The mitochondrial genome of the soybean cyst nematode, *Heterodera glycines*. Genome. 54: 565-574.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity.101:301–320.

Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36:181-184.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology. 59:307-321.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Res. 41: e129.

Handoo Z, Golden AM. 1992. A key and diagnostic compendium to the species of the genus *Hoplolaimus* Daday, 1905 (Nematoda: Hoplolaimidae). J of Nematol.24:45-53.

Holguin, CM, Ma X, Mueller JD, and Agudelo P. 2016. Distribution of *Hoplolaimus* species in soybean fields in South Carolina and North Carolina. Plant Disease. 100: 149-153.

Holroyd N, Sanchez-Flores A. 2012. Producing parasitic helminth reference and draft genomes at the Wellcome Trust Sanger Institute. Parasite Immunol. 34:100-107.

Holterman M, et al. 2006. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. Mol Biol Evol. 23:1792-800.

Howe DK, Denver DR. 2008. Muller's Ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. BMC Evol Biol.8:62.

Hu F, Lin Y, Tang J. 2014. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. BMC Bioinformatics. 15:354.

Humphreys-Pereira DA, Elling AA. 2014. Mitochondrial genomes of *Meloidogyne chitwoodi* and *M. incognita* (Nematoda: Tylenchina): comparative analysis, gene order and phylogenetic relationships with other nematodes. Mol Biochem Parasitol.194: 20-32.

Hunt et al. 2016. The genomic basis of parasitism in the Strongyloides clade of nematodes. Nature Genetics. 48:299-307.

Hyouta et al. 1987. Unusual genetic codes and a novel gene structure for $tRNA^{Ser}_{AGY}$ in starfish mitochondrial DNA. Gene. 56:219-230.

Iwaksaki W, et al.2013. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Mol Biol Evol. 30:2531-2540.

Jacob JE, Vanholme B, Van Leeuwen T, Gheysen G. 2009. A unique genetic code change in the mitochondrial genome of the parasitic nematode *Radopholus similis.* BMC Res Notes. 2:192.

Jenkins WR. 1964. A rapid centrifugal-flotation technique for separating nematodes from soil. Plant Dis. 48:692.

Jühling, et al. 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 40:2833-2845.

Kikuchi T, et al. 2016. Genome-Wide Analyses of Individual *Strongyloides stercoralis* (Nematoda: Rhabditoidea) Provide Insights into Population Structure and Reproductive Life Cycles. PLoS Negl Trop Dis 10: e0005253.

Kim K, Eom KS, Park J. 2006. The complete mitochondrial genome of *Anisakis simplex* (Ascaridida: Nematoda) and phylogenetic implications. Int. J. Parasitol. 36:319-328.

Kim J, et al. 2015. Mitochondrial genomes advance phylogenetic hypotheses for Tylenchina (Nematoda: Chromadorea). Zoologica Scripta. 44:446-462.

Kim J, et al. 2017. Phylogenetic analysis of two Plectus mitochondrial genomes (Nematoda: Plectida) supports a sister group relationship between Plectida and Rhabditida within Chromadorea. Mol Phylogenet Evol. 107:90-102.

Kim T, Kim J, Nadler SA, Park JK. 2016. The complete mitochondrial genome of *Koerneria sudhausi* (Diplogasteromorpha: Nematoda) supports monophyly of Diplogasteromorpha within Rhabditomorpha. Curr Genet. 62:391-403.

Kim T, et al. 2018. The bipartite mitochondrial genome of *Ruizia karukerae* (Rhigonematomorpha, Nematoda). Sci Rep. 8: 7482.

Koenning SR, et al. 1999. Survey of Crop Losses in Response to Phytoparasitic Nematodes in the United States for 1994. J Nematol. 31: 587-618.

Kumar S, Koutsovoulos G, Kaur G, Blaxter M. 2012. Toward 959 nematode genomes. Worm. 1: 42-50.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41:22-28.

Lefort V, Longueville J, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution. 34:2422-2424.

Lavrov DV, Brown WM. 2001. *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. Genetics. 157: 621-637.

Liu GH, et al. 2012. Characterization of the complete mitochondrial genomes of two whipworms *Trichuris ovis* and *Trichuris discolor* (Nematoda: Trichuridae). Infect. Genet. Evol. 12: 1635-1641.

Liu GH et al. 2013. Mitochondrial and nuclear ribosomal DNA evidence supports the existence of a new *Trichuris* species in the endangered François' leaf-monkey. PLoS ONE. 8, e66249.

Lagisz M, Poulin R, Nakagawa S. 2013. You are where you live: parasitic nematode mitochondrial genome size is associated with the thermal environment generated by hosts. Journal of evolutionary biology. 26:683-690.

Leroy S, Duperray C, Morand S. 2003. Flow cytometry for parasite nematode genome size measurement. Mol Biochem Parasitol. 128:91-93.

Lewis SA, Fassuliotis G. 1982. Lance nematodes, *Hoplolaimus* spp., in the Southern United States. In: Riggs RD, editor. Nematology in the Southern Region of the United States. Southern Cooperative Series Bulletin 276. pp. 127-138.

Ma X, Agudelo P, Mulluer JD, Knap HT. 2011. Molecular Characterization and Phylogenetic Analysis of *Hoplolaimus stephanus*. J Nematol. 43:25-34.

van Megen H, et al. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology. 11:927-950.

Molnar R, Bartelmes G, Dinkelacker I, Witte H, Sommer RJ. 2011. Mutation rates and intraspecific divergence of the mitochondrial genome of Pristionchus pacificus. Mol Biol Evol. 28:2317-2326.

Moore WS. 1995. Inferring phylogenies from mtDNA varaiation: mitochondrial-gene trees versus nuclear-gene trees. Evolution. 49:718-726.

Moret BM, Warnow T. 2005. Advances in phylogeny reconstruction from gene order and content data. Methods Enzymol. 395:673-700.

Morley SA, Nielsen BL. 2017. Plant mitochondrial DNA. Front Biosci (Landmark Ed). 22:1023-1032.

Nadler SA, et al. 2007. Molecular phylogeny of clade III nematodes reveals multiple origins of tissue parasitism. Parasitology. 134:1421-1442.

Nagaike T, Suzuki T, Katoh T, Ueda T. 2005. Human mitochondrial mRNAs are stabilized with polyadenylation regulated by mitochondria-specific poly(A) polymerase and polynucleotide phosphorylase. J Biol Chem. 280:19721-19727.

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics. 130:471-498.

Palomares-Rius JE, Cantalapiedra-Navarrete C, Archidona-Yuste A, Blok VC, Castillo P. 2017. Mitochondrial genome diversity in dagger and needle nematodes (Nematoda: Longidoridae). Sci Rep. 7: 41813.

Park JK et al. 2011. Monophyly of clade III nematodes is not supported by phylogenetic analysis of complete mitochondrial genome sequences. BMC Genomics. 12:392.

Phillips WS et al. 2016. The mitochondrial genome of *Globodera ellingtonae* is composed of two circles with segregated gene content and differential copy numbers. BMC Genomics.17:706.

Sagan L. 1967. On the origin of mitosing cells. J Theor Biol. 14:225-247.

Sankoff D, et al. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc Natl Acad Sci U S A. 89:6575-6579.

Shaw KL. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc Natl Acad Sci U S A. 99: 16122-16127.

Sher SA. 1963. Revision of The Hoplolaiminae (Nematoda) II. *Hoplolaimus* Daday, 1905 and *Aorolaimus* N. Gen. Nematologica.9:267–296.

Shi J, Zhang Y, Luo H. Tang J. 2010. Using jackknife to assess the quality of gene order phylogenies. BMC Bioinformatics. 11:168.

Siddiqi MR. 2000. Tylenchida parasites of plants and insects, 2nd edition. Wallingford UK: CABI Publishing. p 270-428.

Siddiqi MR. 1980. The origin and phylogeny of the nematode orders Tylenchida Thorne, 1949 and Aphelenchida n. ord. Helminth. Abstr. Ser. B. 49:143-170.

Singh TR. 2008. Mitochondrial gene rearrangements: new paradigm in the evolutionary biology and systematics. Bioinformation. 3: 95-97.

Springer MS, et al. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol Biol Evol. 18:132-143.

Stajich JE, et al. 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12: 1611-1618.

Stothard P. 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102-1104.

Suga K, Mark Welch DB, Tanaka Y, Sakakura Y, Hagiwara A. 2008. Two circular chromosomes of unequal copy number make up the mitochondrial genome of the rotifer *Brachionus plicatilis*. Mol Biol Evol 25: 1129-1137.

Sultana T, et al. 2013. Comparative analysis of complete mitochondrial genome sequences confirms independent origins of plant-parasitic nematodes. BMC Evol Biol. 13: 12.

Tanifuji G, Archibald JM, Hashimoto T. 2016. Comparative genomics of mitochondria in chlorarachniophyte algae: endosymbiotic gene transfer and organellar genome dynamics. Sci Rep. 6: 21016.

Tillich et al. 2017. GeSeq -- versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45:6-11.

Webb KM, Rosenthal BM. 2011. Next-generation sequencing of the *Trichinella murrelli* mitochondrial genome allows comprehensive comparison of its divergence from the principal agent of human trichinellosis, *Trichinella spiralis*. Infect Genet Evol. 11:116-123.

Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol. 141:173-216.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252-3255.

Young N, et al. 2012. Whole-genome sequence of *Schistosoma haematobium*. Nature Genetics 44: 221-225.

Zhang Y, et al. 2015. The complete mitochondrial genome of *Oxyuris equi*: Comparison with other closely related species and phylogenetic implications. Exp. Parasitol. 159:215-221.

Zhou L, Lin Y, Feng B, Zhao J, Tang J. 2017. Phylogeny analysis from gene-order data with massive duplications. BMC Genomics. 18:760.

Zou H, Jakovlić I, Chen R, Zhang D, Zhang J, Li W-Xand Wang GT. The complete mitochondrial genome of parasitic nematode *Camallanus cotti*: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class. BMC Genomics. 2017;18: 840.

Table 3.1. Tandem repeats finder results. (The bordered indices are in the longest non-coding region.)

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
26--69	18	2.4	19	88	3	63	11	4	6	77	1.11
821--871	23	2.2	23	78	0	57	15	0	9	74	1.06
934--985	19	2.7	20	79	8	54	19	1	5	73	1.14
4054--4109	24	2.4	24	85	14	82	17	0	10	71	1.14
4752--4776	13	1.9	13	100	0	50	36	8	0	56	1.29
4941--4982	17	2.5	17	88	0	57	35	0	2	61	1.09
5626--5659	14	2.5	13	90	9	50	17	0	0	82	0.67
5616--5651	19	1.9	19	88	5	56	13	0	2	83	0.76
5614--5677	21	3	21	80	11	76	15	0	6	78	0.95
5624--5664	13	2.6	17	75	25	54	17	0	0	82	0.66
7648--7686	17	2.3	17	95	0	69	15	0	2	82	0.79
7661--7705	22	2	23	86	4	65	20	2	2	75	1.01
7663--7712	23	2.2	22	82	3	55	22	2	4	72	1.12
7793--7871	6	13.3	6	75	20	53	11	5	2	81	0.96
7796--7858	15	4.2	15	78	13	67	11	3	3	82	0.9
7793--7872	12	6.9	11	75	17	65	11	5	2	81	0.95
7802--7861	20	2.9	20	87	5	75	11	1	3	83	0.84
7825--7901	10	7.7	10	72	5	57	11	2	5	80	0.97
7806--7872	23	2.8	23	80	13	66	10	4	2	82	0.93
8908--8946	9	4.2	9	83	3	51	23	0	0	76	0.78
9682--9715	14	2.5	14	95	4	61	11	0	5	82	0.83
11990--12210	93	2.4	93	88	0	307	29	13	20	36	1.91
12676--12712	14	2.5	15	91	4	58	18	0	5	75	0.99
15076--15139	21	3	21	73	8	58	26	4	1	67	1.19
17188--17287	52	1.9	53	95	2	184	34	2	14	50	1.54
17241--17760	238	2.2	233	93	3	873	29	4	13	53	1.58
17241--17761	235	2.2	237	93	4	885	28	4	13	53	1.58
17776--18007	45	5.1	45	92	2	376	37	2	12	46	1.56
17776--18013	91	2.6	91	93	2	399	38	2	13	46	1.56
17949--18061	34	3.4	34	68	16	101	34	1	15	48	1.55
17947--18070	65	1.9	63	93	3	212	34	1	15	48	1.55
17903--18115	110	2	108	92	3	356	36	1	13	47	1.54
18012--18138	44	2.9	44	92	5	220	37	2	13	46	1.56
17856--18138	154	1.8	155	93	3	498	37	2	13	46	1.55
19434--19573	49	2.9	47	77	11	140	20	6	10	62	1.49
21000--21165	43	3.8	43	93	2	287	34	18	21	25	1.96
21000--21165	87	1.9	87	93	2	289	34	18	21	25	1.96
23756--23792	18	2.2	17	86	13	51	27	0	0	72	0.84
24166--24225	21	2.8	20	80	7	57	16	1	6	75	1.1

Table 3.2. Annotation results of two different genetic codes (discrepancies are highlighted, Bordered regions indicate overlaps.)

Genetic code 5						Genetic code 14					
Name	Start	Stop	Strand	Length	Codon	Name	Start	Stop	Strand	Length	Codon
trnK(aaa)	1115	1183	+	69							
						trnN(aaa)	1115	1183	+	69	
rrnS	2055	2124	+	70		rrnS	2055	2124	+	70	
trnS2(tca)	2130	2192	+	63		trnS2(tca)	2130	2192	+	63	
trnY(tac)	2192	2247	+	56							
trnW(tga)	2248	2308	+	61		trnW(tga)	2248	2308	+	61	
						nad4l	2293	2412	+	120	
trnW(tgg)	2314	2373	+	60							
nad1	2393	3223	+	831	ATT/ATG	nad1	2393	3223	+	831	ATT/ATG
trnQ(caa)	3235	3284	+	50		trnQ(caa)	3235	3284	+	50	
trnC(tgc)	3344	3416	+	73		trnC(tgc)	3344	3416	+	73	
trnE(gaa)	3418	3473	+	56		trnE(gaa)	3418	3473	+	56	
trnS1(aga)	3526	3583	+	58		trnS1(aga)	3526	3583	+	58	
trnL2(tta)	3584	3639	+	56		trnL2(tta)	3584	3639	+	56	
trnV(gta)	3676	3729	+	54		trnV(gta)	3676	3729	+	54	
trnL1(cta)	3907	3962	+	56		trnL1(cta)	3907	3962	+	56	
cox2	4053	4625	+	573	ATT/CTA	cox2	4047	4625	+	579	GTT/CTA
trnH(cac)	4636	4687	+	52		trnH(cac)	4636	4687	+	52	
rrnL	5047	5575	+	529		rrnL	5047	5575	+	529	
nad3	5685	5915	+	231	ATA/TGG	nad3	5670	5915	+	246	
cox1	6068	7480	+	1413	AGC/GCA	cox1	6068	7480	+	1413	AGC/GCA
atp6	7999	8229	+	231	ATT/GAA	atp6	7999	8229	+	231	ATT/GAA
						trnM(atg)	8235	8287	+	53	
nad5	8336	9418	+	1083	ATA/AAT	nad5	8354	9409	+	1056	ATA/AAT
trnD(gac)	9829	9898	+	70		trnD(gac)	9829	9898	+	70	
nad2	9878	10606	+	729	ATT/TTT	nad2	9875	10645	+	771	TTA/AAA
trnI(atc)	10755	10810	+	56							
cox3	10809	11561	+	753	TTG/TGA	cox3	10821	11561	+	741	TAT/TGA
trnF(ttc)	11981	12049	+	69		trnF(ttc)	11981	12049	+	69	
						trnY(taa)	12246	12299	+	54	
nad4	12404	13360	+	957	TTG/TGA	nad4	12422	13375	+	954	TTA/TTT
trnN(att)	13526	13579	+	54		trnN(aat)	13526	13579	+	54	
cob	13510	14625	+	1116	ATT/ATT	cob	13612	14625	+	1014	ATG/ATT
trnP(cca)	14659	14712	+	54		trnP(cca)	14659	14712	+	54	
nad6	14755	15159	+	405	ATT/TTA						
trnG(gga)	15163	15219	+	57		trnG(gga)	15163	15219	+	57	
						nad6	15435	15806	-	372	GAT/GTT
nad4l	15273	15428	+	156	ATT/GTC						
trnR(cga)	23114	23185	-	72		trnR(cga)	23114	23185	-	72	
						trnK(aag)	24043	24097	+	55	

Table 3.3. Codon usage (discrepancies are highlighted.)

Code 5					Code 14					Code 5					Code 14				
AA	**Codon**	**#**	**/1000**	**Fraction**	**AA**	**Codon**	**#**	**/1000**	**Fraction**	**AA**	**Codon**	**#**	**/1000**	**Fraction**	**AA**	**Codon**	**#**	**/1000**	**Fraction**
Ala	GCG	4	1.12	0.14	Ala	GCG	4	1.12	0.14	Pro	CCG	3	0.84	0.08	Pro	CCG	3	0.84	0.08
Ala	GCA	5	1.4	0.18	Ala	GCA	5	1.4	0.18	Pro	CCA	8	2.24	0.22	Pro	CCA	8	2.24	0.22
Ala	GCT	15	4.2	0.54	Ala	GCT	15	4.2	0.54	Pro	CCT	23	6.43	0.62	Pro	CCT	23	6.43	0.62
Ala	GCC	4	1.12	0.14	Ala	GCC	4	1.12	0.14	Pro	CCC	3	0.84	0.08	Pro	CCC	3	0.84	0.08
Cys	TGT	42	11.75	0.7	Cys	TGT	42	11.75	0.7	Gln	CAG	15	4.2	0.47	Gln	CAG	15	4.2	0.47
Cys	TGC	18	5.03	0.3	Cys	TGC	18	5.03	0.3	Gln	CAA	17	4.76	0.53	Gln	CAA	17	4.76	0.53
Asp	GAT	77	21.54	0.89	Asp	GAT	77	21.54	0.89	Arg	CGG	9	2.52	0.69	Arg	CGG	9	2.52	0.69
Asp	GAC	10	2.8	0.11	Asp	GAC	10	2.8	0.11	Arg	CGA	1	0.28	0.08	Arg	CGA	1	0.28	0.08
Glu	GAG	49	13.71	0.47	Glu	GAG	49	13.71	0.47	Arg	CGT	3	0.84	0.23	Arg	CGT	3	0.84	0.23
Glu	GAA	55	15.38	0.53	Glu	GAA	55	15.38	0.53	Arg	CGC	0	0	0	Arg	CGC	0	0	0
Phe	TTT	623	174.27	0.9	Phe	TTT	623	174.27	0.9	Ser	AGG	29	8.11	0.13	Ser	AGG	29	8.11	0.13
Phe	TTC	70	19.58	0.1	Phe	TTC	70	19.58	0.1	Ser	AGA	38	10.63	0.18	Ser	AGA	38	10.63	0.18
Gly	GGG	41	11.47	0.28	Gly	GGG	41	11.47	0.28	Ser	AGT	46	12.87	0.21	Ser	AGT	46	12.87	0.21
Gly	GGA	29	8.11	0.2	Gly	GGA	29	8.11	0.2	Ser	AGC	7	1.96	0.03	Ser	AGC	7	1.96	0.03
Gly	GGT	63	17.62	0.44	Gly	GGT	63	17.62	0.44	Ser	TCG	10	2.8	0.05	Ser	TCG	10	2.8	0.05
Gly	GGC	11	3.08	0.08	Gly	GGC	11	3.08	0.08	Ser	TCA	23	6.43	0.11	Ser	TCA	23	6.43	0.11
His	CAT	26	7.27	0.81	His	CAT	26	7.27	0.81	Ser	TCT	51	14.27	0.24	Ser	TCT	51	14.27	0.24
His	CAC	6	1.68	0.19	His	CAC	6	1.68	0.19	Ser	TCC	12	3.36	0.06	Ser	TCC	12	3.36	0.06
					Ile	ATA	82	22.94	0.24	Thr	ACG	5	1.4	0.13	Thr	ACG	5	1.4	0.13
Ile	ATT	234	65.45	0.9	Ile	ATT	234	65.45	0.68	Thr	ACA	6	1.68	0.15	Thr	ACA	6	1.68	0.15
Ile	ATC	26	7.27	0.1	Ile	ATC	26	7.27	0.08	Thr	ACT	19	5.31	0.47	Thr	ACT	19	5.31	0.47
Lys	AAG	43	12.03	0.41	Lys	AAG	43	12.03	1	Thr	ACC	10	2.8	0.25	Thr	ACC	10	2.8	0.25
Lys	AAA	61	17.06	0.59						Val	GTG	28	7.83	0.15	Val	GTG	28	7.83	0.15
Leu	TTG	124	34.69	0.22	Leu	TTG	124	34.69	0.22	Val	GTA	36	10.07	0.2	Val	GTA	36	10.07	0.2
Leu	TTA	279	78.04	0.49	Leu	TTA	279	78.04	0.49	Val	GTT	103	28.81	0.57	Val	GTT	103	28.81	0.57
Leu	CTG	18	5.03	0.03	Leu	CTG	18	5.03	0.03	Val	GTC	14	3.92	0.08	Val	GTC	14	3.92	0.08
Leu	CTA	49	13.71	0.09	Leu	CTA	49	13.71	0.09	Trp	TGG	73	20.42	0.5	Trp	TGG	73	20.42	0.5
Leu	CTT	81	22.66	0.14	Leu	CTT	81	22.66	0.14	Trp	TGA	73	20.42	0.5	Trp	TGA	73	20.42	0.5
Leu	CTC	13	3.64	0.02	Leu	CTC	13	3.64	0.02						Tyr	TAA	228	63.78	0.48
Met	ATG	53	14.83	0.39	Met	ATG	53	14.83	1	Tyr	TAT	220	61.54	0.88	Tyr	TAT	220	61.54	0.46
Met	ATA	82	22.94	0.61						Tyr	TAC	29	8.11	0.12	Tyr	TAC	29	8.11	0.06
					Asn	AAA	61	17.06	0.39	End	TAG	126	35.24	0.36	End	TAG	126	35.24	1
Asn	AAT	81	22.66	0.84	Asn	AAT	81	22.66	0.52	End	TAA	228	63.78	0.64					
Asn	AAC	15	4.2	0.16	Asn	AAC	15	4.2	0.1										

Table 3.4. Species used for phylogenetic analyses and protein-coding gene order survey in this study

Species	Accessions	Sizes (bp)	Species	Accessions	Sizes (bp)
Acanthocheilonema viteae	NC_016197	13,724	*Nippostrongylus brasiliensis*	NC_033886.1	13,355
Aelurostrongylus abstrusus	NC_019571.1	13,913	*Oesophagostomum dentatum*	GQ888716.1	13,869
Ancylostoma ceylanicum	NC_035142.1	13,660	*Oesophagostomum quadrispinulatum*	NC_014181.1	13,681
Ancylostoma tubaeforme	NC_034289.1	13,730	*Onchocerca flexuosa*	NC_016172.1	13,672
Agamermis.spBH2006	NC_008231	16,561	*Onchocerca ochengi*	KX181290.2	13,744
Angiostrongylus cantonensis	GQ398121.1	13,497	*Oscheius chongmingensis*	KP257594.1	15,413
Angiostrongylus costaricensis	NC_013067.1	13,585	*Parafilaroides normani*	NC_024656.1	13,414
Angiostrongylus malaysiensis	NC_030332.1	13,516	*Paralongidorus litoralis*	NC_033868.1	12,763
Angiostrongylus vasorum	NC_018602.1	13,422	*Parascaris univalens*	KM216010.1	14,350
Anisakis pegreffii	LC222461.1	14,002	*Passalurus ambiguus*	NC_028345.1	14,023
Aphelenchoides besseyi	KJ739799.1	16,216	*Philometroides sanguineus*	NC_024931.1	14,378
Ascaris lumbricoides	HQ704900.1	14,303	*Pratylenchus vulnus*	NC_020434.1	21,656
Ascaris suum	HQ704901.1	14,311	*Pristionchus pacificus*	JF414117.1	15,954
Baylisascaris procyonis	NC_016200.1	14,781	*Protostrongylus rufescens*	NC_023262.1	13,619
Brugia malayi	NC_004298.1	13,657	*Pseudoterranova azarasi*	NC_027163.1	13,954
Bursaphelenchus xylophilus	JQ514068.1	12,945	*Radopholus similis*	FN313571.1	16,791
Caenorhabditis briggsae	AC186293.1	14,420	*Rhigonema thysanophora*	NC_024020.1	15,015
Caenorhabditis elegans	NC_001328.1	13,794	*Romanomermis culicivorax*	NC_008640.1	26,194
Caenorhabditis nigoni	KP259621.2	13,856	*Romanomermis iyengari*	EF175764.1	18,919
Caenorhabditis remanei	KR709159.1	13,977	*Romanomermis nielseni*	EF175763	15,546
Caenorhabditis sinica	EU407780.1	13,537	*Rotylenchulus reniformis*	CM003310.1	24,572
Caenorhabditis tropicalis	NC_025756.1	13,874	*Setaria digitata*	KY284626.1	13,814
Contracaecum rudolphii	FJ905109.1	14,022	*Spirocerca lupi*	NC_021135.1	13,780
Cooperia oncophora	NC_004806.1	13,636	*Steinernema carpocapsae*	NC_005941.1	13,925
Cucullanus robustus	GQ332426.1	13,972	*Strelkovimermis spiculatus*	NC_008047.1	18,030
Cylicocyclus nassatus	NC_032299.1	13,846	*Strongyloides papillosus*	LC050210.1	13,909
Dirofilaria immitis	NC_005305.1	13,814	*Strongyloides ratti*	NC_028623.1	16,609
Enterobius vermicularis	EU281143.1	14,010	*Strongyloides stercoralis*	NC_028624.1	13,751
Globodera ellingtonae (I)	KU726971.1	17,757	*Strongyloides venezuelensis*	LC050213.1	15,956
Globodera ellingtonae (II)	KU726972.1	14,365	*Strongylus equinus*	NC_026868.1	14,545
Gnathostoma spinigerum	NC_027726.1	14,079	*Thelazia callipaeda*	NC_018363.1	13,668
Gongylonema pulchrum	NC_026687.1	13,798	*Toxascaris leonina*	NC_023504.1	14,310
Haemonchus contortus	NC_010383.2	14,055	*Toxocara canis*	NC_010690.1	14,322
Heliconema longissimum	NC_016127.1	13,610	*Toxocara cati*	NC_010773.1	14,029
Heterodera glycines	HM640930.1	14,915	*Toxocara malaysiensis*	NC_010527.1	14,266
Heterorhabditis bacteriophora	NC_008534.1	18,128	*Trichinella spiralis*	KM357422.1	16,584
Hexamermis agrotis	EF368011.1	24,606	*Trichostrongylus axei*	NC_013824.1	13,653
Hoplolaimus columbus	MH657221	25,228	*Trichuris sp.GHL2013*	KC461179.1	14,147
Koerneria sudhausi	NC_029233.1	16,005	*Trichuris suis*	NC_017747.1	14,436
Loa loa	HQ186250.1	13,590	*Trichuris trichiura*	NC_017750.1	14,046
Longidorus vineacola	NC_033867.1	13,519	*Triodontophorus brevicauda*	NC_026729.1	14,305
Meloidogyne arenaria	NC_026554.1	17,580	*Wellcomia siamensis*	NC_016129.1	14,128
Meloidogyne chitwood	KJ476150.1	18,201	*Wuchereria bancrofti*	HQ184469.1	13,636
Meloidogyne enterolobii	KP202351.1	17,053	*Xiphinema americanum*	AY382608.1	12,626
Meloidogyne graminicola	KJ139963.1	19,589	*Xiphinema pachtaicum*	NC_033870.1	12,489
Meloidogyne incognita	KJ476151.1	17,662	*Xiphinema rivesi*	NC_033869.1	12,624
Meloidogyne javanica	NC_026556.1	18,291			

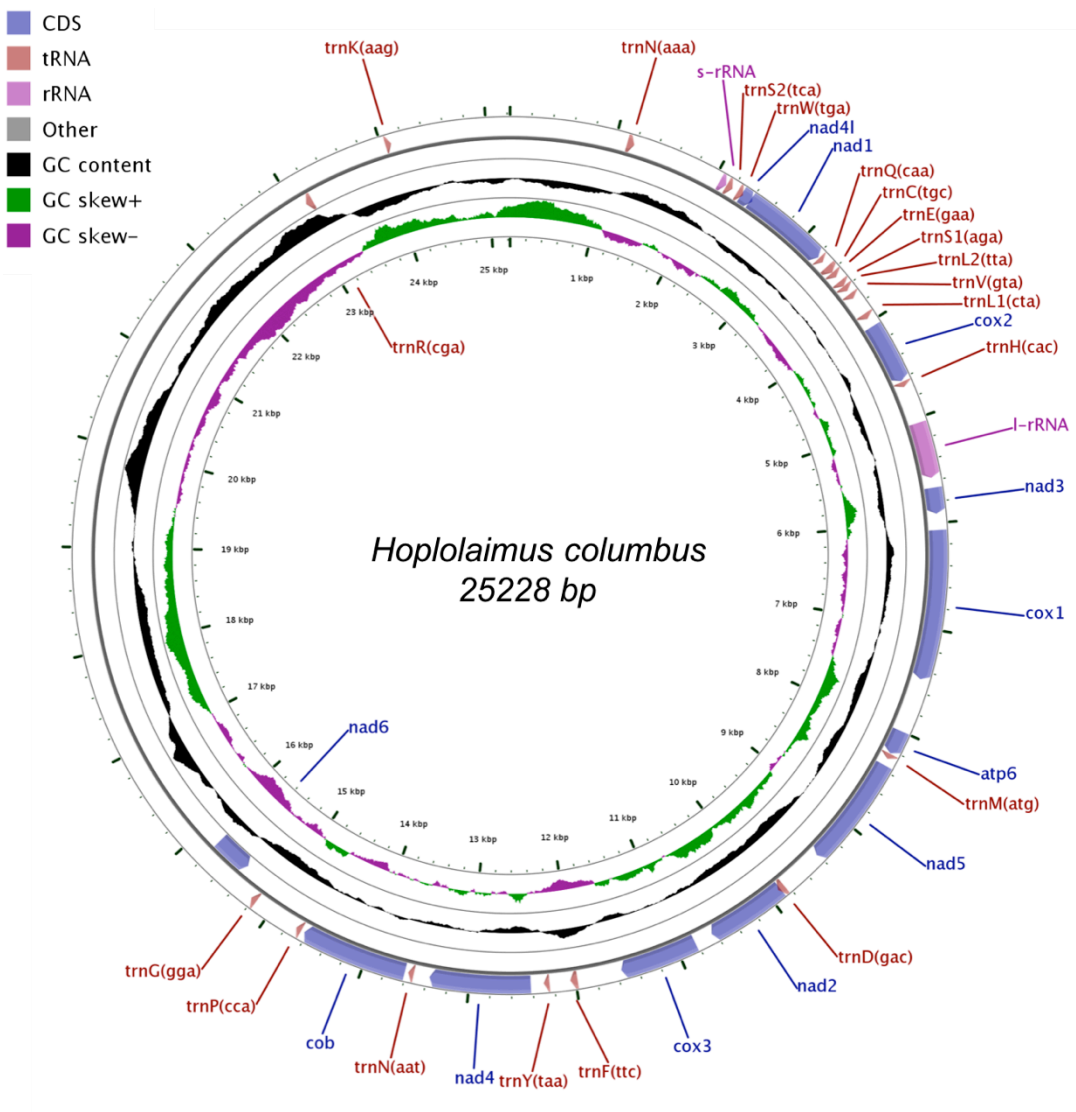

Figure 3.1. ***Hoplolaimus columbus*** **mitochondrial genome map using mitochondrial genetic code 14**

```
CLUSTAL format alignment by MAFFT (v7.408)
nd4L-5          atttttattataatttttttagaaataacaggtttttgattttctttagttttttgatt
nd4L-14         ggttt-------------------atcccaggtttttattagtctaatagtttattaaga
                . ***                   **  *********     *..  ****** **.*
nd4L-5          aatttttctttttaattc--aattttcttatttttaggattaattgtgattactagggtct
nd4L-14         atttttgtttccaaaacaaaaggtattttattttagatttattagagatttttggttgat
                * **** .**..**  *  **  * .**  ******. *** * * **** .*.*    *
nd4L-5          atggtttttttggtttttgtttttttatttttaaaaagagt
nd4L-14         tgttttatttagtttttgt------------------
                    ** ***.********

CLUSTAL format alignment by MAFFT (v7.408)
nd6-5           atttttttaatttttttattattatttttttggtggtctgagagtctctgatccattaaaa
nd6-14          g---ataaagtatatttttagttatttttatggacagaaggagagacagagat---ttaagg
                .    *  *.* * *** * .******** * .    ..*** *   ***   ****..
nd6-5           ataggtttattaattatttttagattattattattaggaatttatttaataaattggtta
nd6-14          gtagatttatctttcatttttatttttttttaaaaaaaaaggtaaataggt----------
                .***.*****.  *.*******  ** **   *  *..*. * * * ..*
nd6-5           ggatttctaataatttttcttgatttatataatttttattagtggactatttttaatttta
nd6-14          ----------tatttttattatttcttataggcttagtta---gacaaatttataattat
                          ** **** **. **. ****. .** .***   *** * ***  * **
nd6-5           gtctattgtactagtttgagttttgaaaaaaaaggttatttctctattttaaggttattt
nd6-14          ttttcaggtaataatttgg---tggccagggatcttgggttttctgctgtaagaaatatt
                 *.*   *** **.****.   * *  *...*   * . **.***..* ****.    **
nd6-5           tttttaatttttagtcttaagttaaatttttattttagaagaagttttattaaggtttta
nd6-14          tctattgtgccctttacctacgtcttacctatagatataggtttttctgccattagtaatg
                *.* * .* ..*   *.** **.  *..* **  *  ...    *.*  .   .**  *.
nd6-5           ttaatattgaattattttaatttaatattaatt--attatttttttta--tgcttttatt
nd6-14          tttctattgaacc-taatagtttatcattaattccatccttatctttagttggttctact
                **  *******.. *  **.**** .*******  **. ** *.****  ** **.**.*
nd6-5           attatttcttaatattatattatttttttcaaggggctctccgatcctta
nd6-14          tctcttcctccgtttaatgtgattattcctaagaacttgagagtt----
                 .* **.**. .* * **.* *** **.* *.*..**.   ..*.

CLUSTAL format alignment by MAFFT (v7.408)
nd6-5           -attttttttaatttttttattattatttttttggtggtctgagagtctctgatccattaaa
nd6-14-reverse  aactctcaagttcttaggaataatcacattaaacggaggaagagaagtagaaccaactaa
                 *.*.*.  . *.**   * ** *  . ** ...**   .****   . ** *** . **
nd6-5           aataggtttattaattatttttagattattattattaggaatttatttaataaattggtt
nd6-14-reverse  agataaggatggaattaatgataaactattaggttcaatagaaacattactaatggcaga
                *.  ..      ***** *  **.*.*****   *.*. *.     *** ***    .
nd6-5           aggatttctaataatttttcttgatttatata----------------atttttattagtg
nd6-14-reverse  aaaacctatatctataggtaagacgtaggtaaaggcacaatagaaatatttcttacagca
                *..*..* ** . **    .  **. ** .**                ****.*  .**..
nd6-5           gactatttttaattttagtctattgtactagtttgagttttgaaaaaaaaggttatttct
nd6-14-reverse  gaaaacccaagatccctggccaccaaattatt----acctgaaaaataattataaatttg
                **  *...  .**... * *.*... *.** *    ...* .**** **  .* * **.
nd6-5           ctattttaaggttatttttttttaatttttagtcttaagttaaatttttatttttagaagaa
nd6-14-reverse  tctaactaagcctataagaaataataaaaataacctatttacctttttttttaaaaaaaa
                ..   .**** .***      ****    *   .. * ***  ***** *** *.**.**
nd6-5           gttttattaaggttttattaatattgaa----ttattttaatttaatattaattattatt
nd6-14-reverse  taaaaatgaaagataaatctacccttaaatctctgtctccttctgtccataaaaataact
                     ** **.* *  **. *. .* **    .*.*.*.  *.*. .  ***  ** *.*
nd6-5           ttttttatgcttttattattatttcttaatattatattattttttcaaggggctctccga
nd6-14-reverse  aaaaatatacttt-------------------------------------------------
                     ***.****
nd6-5           tcctta
nd6-14-reverse  ---atc
                    *
```

Figure 3.2. Alignment of nad4l and nad6 genes from two annotation results, asterisks indicate positions of identical nucleotides in the sequences.

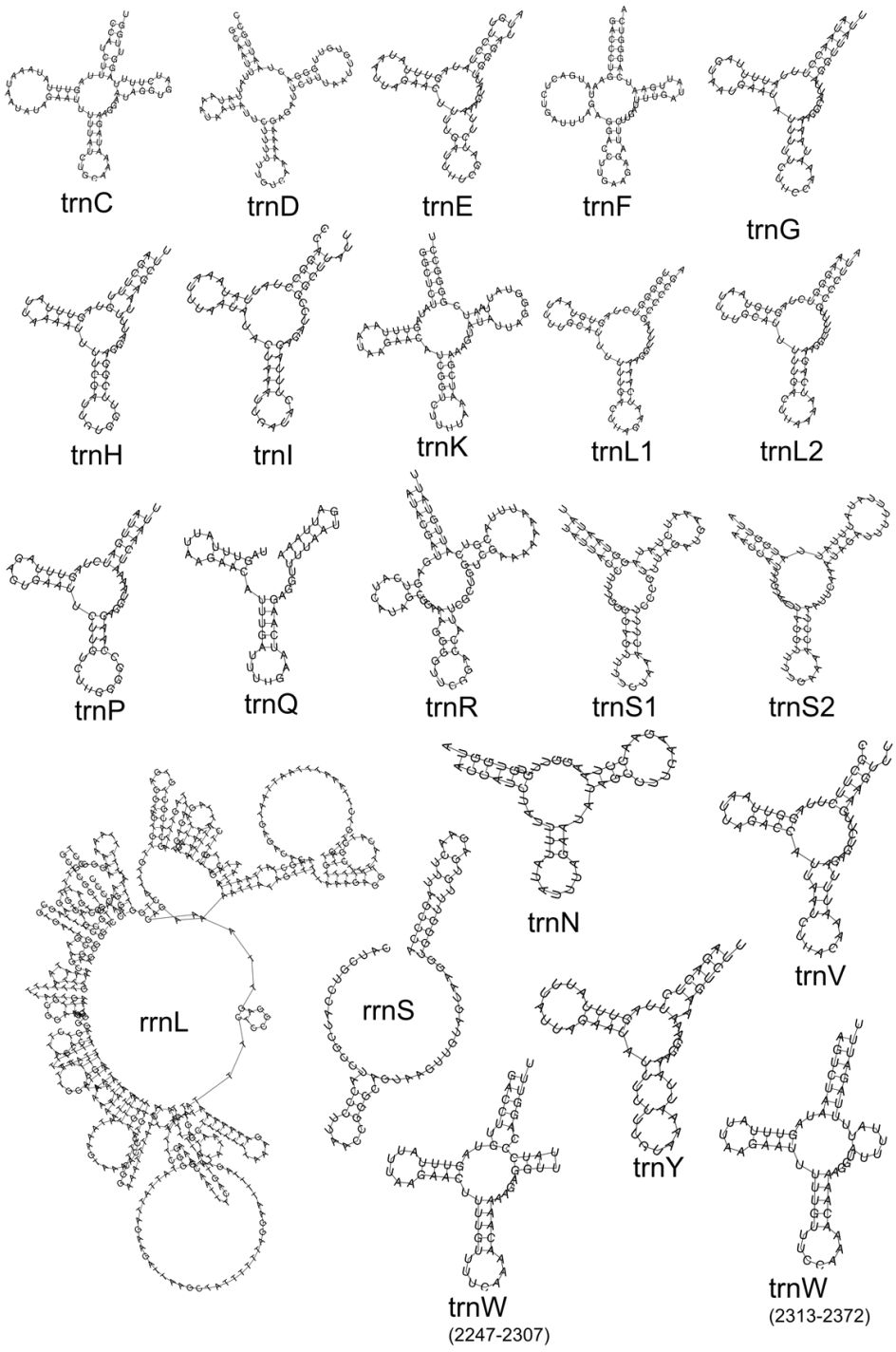

Figure 3.3. RNA structure predictions based on genetic code 5, rotated vertically by anticodon loop

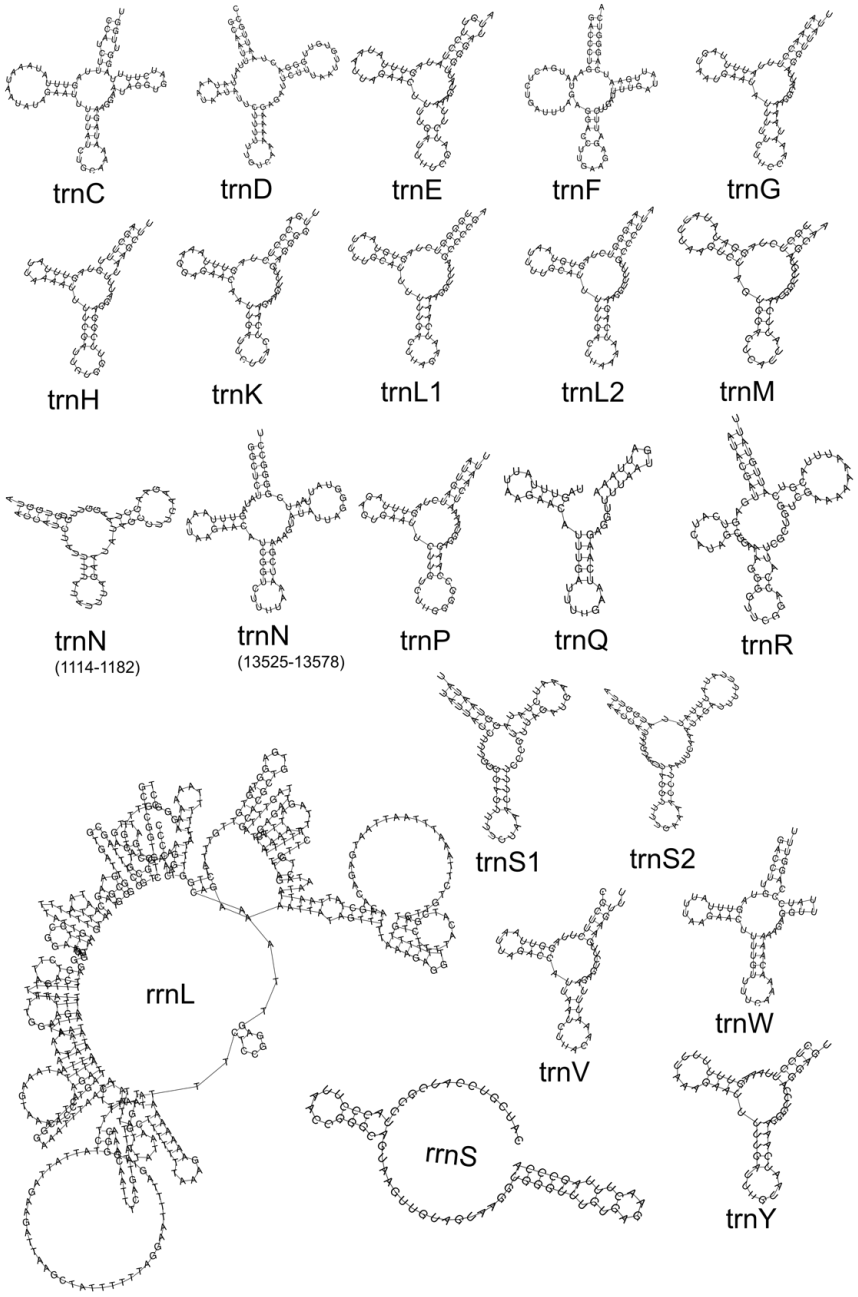

Figure 3.4. RNA structure predictions based on genetic code 14, rotated vertically by anticodon loop

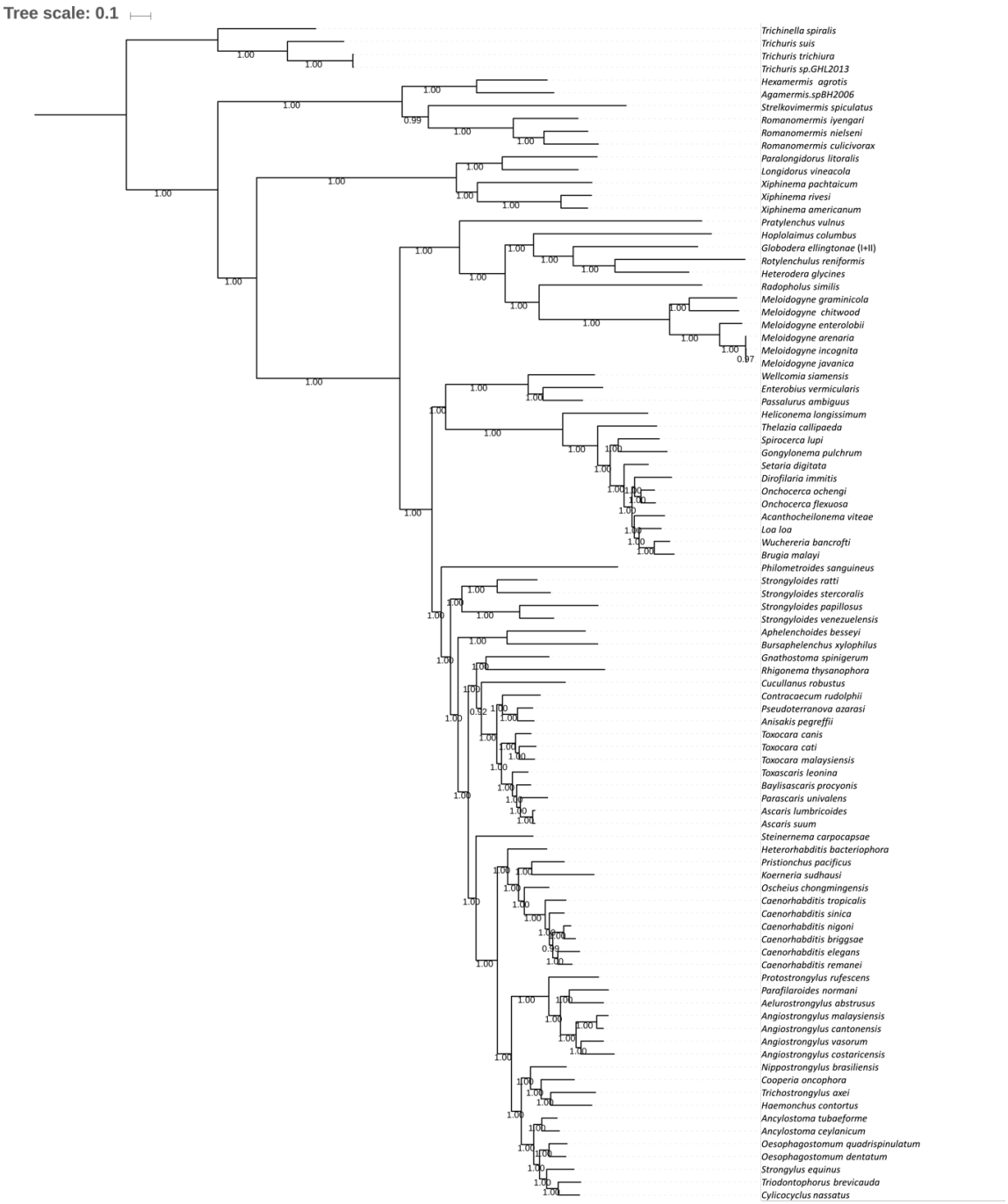

Figure 3.5. Phylogenetic relationships (identical Maximum Likelihood results from both aLRT-SH-like and aBayes methods) of 92 nematode species using concatenated sequences of 12 protein-coding genes

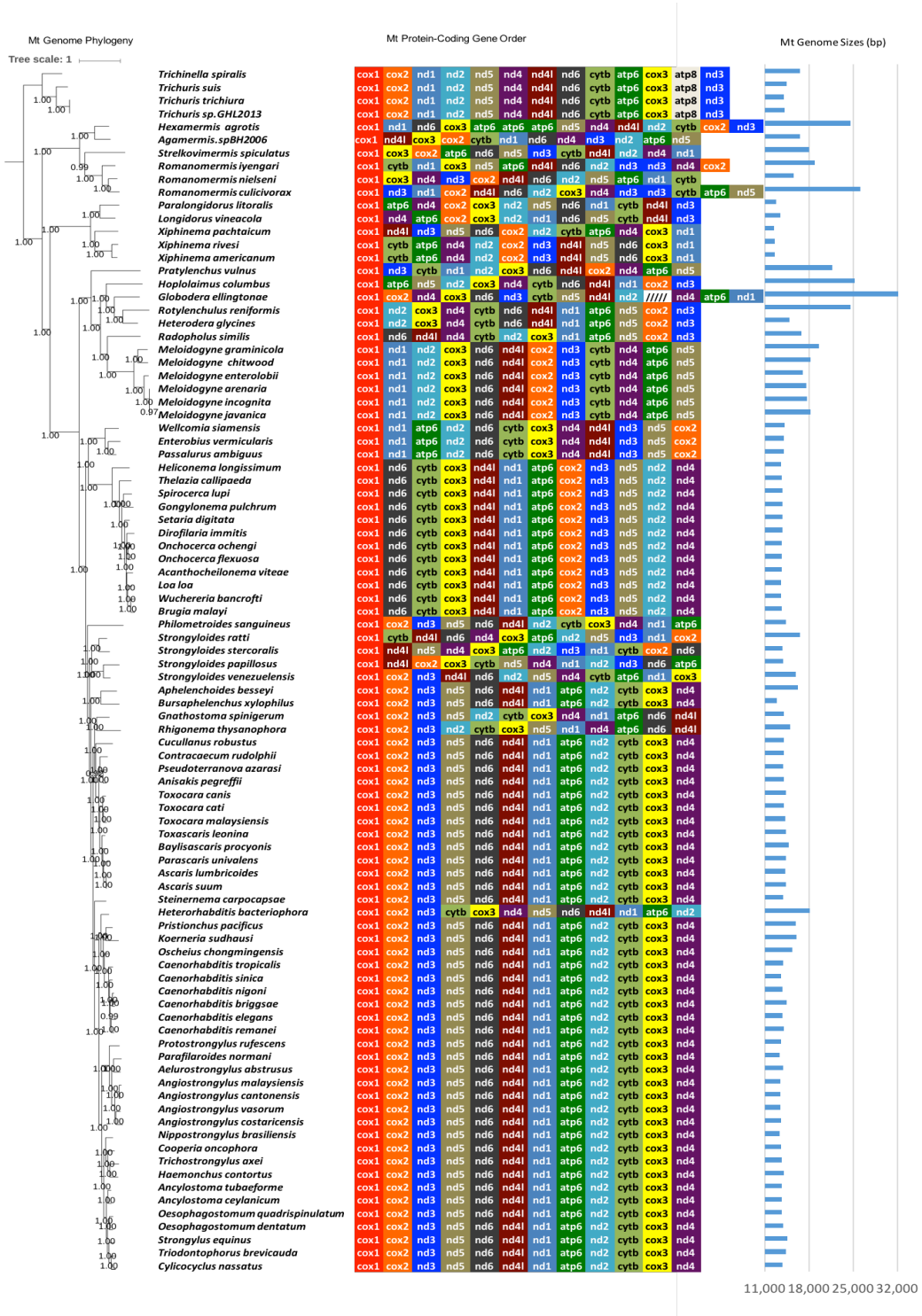

Figure 3.6. Protein-coding gene order arrangement and mitochondrial genome sizes aligning with phylogenetic relationships

CHAPTER IV.

PHYLOGENETIC INFORMATIVENESS INVESTIGATION OF MITOCHONDRIAL PROTEIN-CODING GENES IN NEMATODA

Introduction

The Nematoda Phylum is one of the largest animal group with high diversity on both biological behavior and morphological characters on the Earth. They could be found parasitic in human, mammals, birds, plants, insects, or free-living in freshwater, oceans, soil, and even in the ice (Platt, 1994; Bongers and Ferris, 1999; Saint Andre et al., 2002; Raymond et al., 2016). The early taxonomic classifications of the Nematoda went through over 60 years based on anatomical and morphological characters. Later, molecular phylogenetic analyses in nematode researches have contributed invaluable efforts on reconstructing evolutionary relationships. Nuclear gene markers, such as small-subunit ribosomal DNA and internal transcribe space sequences, have been implemented widely to unravel the classification system of nematodes (Blaxter et al., 1998; De Ley and Blaxter, 2002; Holterman et al., 2006; van Megen et al., 2009). Nuclear markers were also widely useful on species identification and population diagnosis (Ma et al., 2011; Powers et al., 1997).

Mitochondrial DNA (mtDNA) is another significant source of phylogenetic markers for the nematode phylogeny. The mitochondrion is an organelle widely found in metazoan, which is thought to be descended about 1 billion years ago from the alpha-

proteobacteria (Sagan, 1967; Andersson et al., 2003; Boussau et al., 2004). It normally contains a single circular molecule of DNA, while multiple circular mitochondrial molecules have been reported (Armstrong et al., 2000; Suga et al., 2008; Philips et al., 2016). Because they are relatively easier to manipulate, clonally inherited, single-copy and abundant (Simon et al., 1994), mtDNA markers have been widely used for barcoding in many groups of organisms to identify new nematode species(Kim et al., 2017), elucidate the population structure (Cabasan et al., 2018) and inferring the evolutionary histories of animal species (Laetsch et al., 2012).

Systematic studies have shown that some molecular markers are better than others for reconstructing evolutionary relationships among taxa at particular level of divergence (Simon et al., 1994), and different regions of mtDNA present diverse phylogenetic signal in population studies (Cabasan et al., 2018). Therefore, it is predictable that phylogenetic signals are not averagely distributed on mitochondrial genome, though their physical proximity. A metazoan mitochondrial genome is usually a single compact circular DNA with a conserved gene content. There are normally 12 or 13 protein-coding genes (PCGs) which including ATPase complex genes (atp6 and/or atp8 gene), cytochrome b gene (cytb gene), subunits I-III of cytochrome c oxidase (cox1, cox2, and cox3 genes), and 7 subunits of the respiratory chain NADH dehydrogenase (nad1, nad2, nad3, nad4, nad4l, nad5, and nad6 genes) (Chomyn and Attardi, 1987) and some other genes coding tRNAs and rRNAs, However, most of the reported nematode mitochondrial genomes are missing the *atp8* gene except *Trichinella spiralis* (Lavrov and Brown, 2001) and

Trichuris spp. (Liu et al., 2012) in Trichinellida order. Moreover, pseudogenes are reported in *Caenorhabditis briggsae* and *Camallanus cotti* (Howe and Denver, 2008; Zou et al., 2017), and repeated protein-coding genes are present in the mitochondrial genomes of *Hexamermis agrotis*, *Romanomermis culicivorax* and *R. iyengari* (Beck-Azevedo and Hyman, 1993; Lagisz et al., 2013). Although there are 12 PCG candidates, the most popular phylogenetic markers is the cox1 gene and the cytb gene. The phylogenetic analyses based on the cox1 marker normally present higher robust results than nuclear markers. However, it has been reported that cox1 gene may show lower diversity than other genes in population studies (Cabasan et al., 2018). Therefore, choosing a suitable phylogenetic marker is an important step for nematode phylogeny. Another method to increase the analytic robust of phylogeny is to add multiple markers together to make a "super gene" as one marker. It has become prevalent especially in mitochondrial phylogeny since mitochondrial genomes are relatively easier to obtain (Park et al., 2011; Sultana et al 2013; Kim et al. 2017).

In our study, to better understand the utility of mitochondrial protein-coding genes in the nematode phylogenetic analyses, we investigated the phylogenetic signal from 93 nematode mitochondrial genomes including two recently sequenced plant-parasitic nematode species (*Hoplolaimus columbus* and *H. galeatus*). Two methods were utilized to check the signal: one is to compare phylogenetic results from different markers based on their topological distance (Robinson-Foulds symmetric difference) and similarity (frequency of edges in the reference tree found in target) (Huerta-Cepas, et al., 2016); the

other is to follow the conception of phylogenetic informativeness (PI) (Townsend, 2007) and to profile the PI value of each phylogenetic marker. We diagnosed the phylogenetic signal of 12 common protein-coding genes that are widely found in nematodes, and also tested the phylogenetic signal of concatenated-genes markers of diverse concatenation patterns.

Materials and methods

Sequence data

Mitochondrial genome data utilized in this investigation were of 93 nematode species, listed in the Table 4.1. Besides directly downloading from GenBank, we *de novo* sequenced 2 lance nematode species (*Hoplolaimus* spp.) using Whole Genome Amplification (WGA) and Illumina MiSeq technique. The live nematode specimens were merged in distilled water and left starving over two weeks. Before DNA extraction, nematodes were merged in hydrogen peroxide 3% solution (Aaron Industry) over 5 mins, then washed in distilled water three times to remove potential microorganism contamination on the surface. Finally, nematodes were merged in DNA Away solution (Molecular BioProducts, Inc.) to remove potential DNA and DNase contamination and washed using PCR-grade water three times. WGA was performed for obtaining adequate total DNA of individual nematode using Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare). Three replications were performed and the one with the best DNA concentration tested by Qubit fluorometer (Invitrogen) was selected for

MiSeq library preparation. The library preparation followed the protocol of Nextera XT kit (Illumina). Library concentration and fragment size were respectively determined using Qubit fluorometer (Invitrogen) and Bioanalyzer 2100 (Agilent Technologies). Sequencing processes used Illunima MiSeq v3 kit, and a number of 56409068 in total of 300bp-reads were yielded. A reliable result of 98.1067% of total reads is recognized as high quality reads passed through Illumina internal filter, and approximately 13 Gb sequence data have quality scores (Q-score) higher than 30.

Mitochondrial genome assembly was performed using NovoPlasty2.7.2 (Dierckxsens, 2016), MIRA (Chevreux et al., 1999) and MITObim (Hahn et al., 2013). Annotation results were predicted by MITOS online webserver (Bernt et al., 2013) using genetic code 14.

Protein coding genes alignment and formatting files

All PCG information of 93 nematodes were reorganized using BioPerl (Stajich et al., 2002) and BioPython (Cock et al., 2009) to be 12 fasta files for 12 PCGs. The alignments were first performed on MAFFT version 7 webserver (Kuraku, 2013) and saved as fasta files, then converted into phylip files and nexus files. Sequences on minus strand were diagnosed by MAFFT during alignment and were modified using Molecular Biology Online Apps (http://www.molbiotools.com/).

To concatenate sequences, we used SequenceMatrix (Vaidya et al., 2011) to deal with 12 aligned fasta files. There are four different strategies to concatenate sequences for different investigation purposes: (1) concatenated all fasta files of 12 PCGs; (2) concatenated only 11 genes, and created 12 files of different combinations; (3) concatenated genes from 2 genes to 10 genes to create another 9 files, according to the gene size from the minimum to the largest; (4) concatenated genes from 2 genes to 10 genes to create another 9 files, according to the gene size from the largest to the smallest. There were 43 alignments finally for phylogenetic analyses.

Then alignment files were converted into phylip and nexus format on webserver Converter of Phylogeny.fr (Dereeper and Guignon et al., 2008; Dereeper et al., 2010).

Compare phylogenetic trees and calculate distances

Phylogenetic analyses of all 43 alignments were performed using both Maximum Likelihood (ML) and Bayesian inference (BI) methods. ML analyses were performed on PhyML3.0 webserver (Guindon et al., 2010). The number of substitution rate categories was set as 4, and the final model for ML analysis was GTR+G+I, selected by SMS (Smart Model Selection in PhyML) (Lefort et al., 2017). BI analyses were performed using MrBayes3.2.6 (Ronquist et al., 2012; Huelsenbeck, 2001). The evolutionary model was set as GTR+G+I, 1000000 generations of Markov chain Monte Carlo analyses, sumt burnin set as 10000, and sump burnin set as 2.

Phylogenetic tree comparison were performed using the Environment for Tree Exploration v3 (ETE 3) (Huerta-Cepas et al., 2016). The 12-gene-concatenated tree from both methods were set as reference trees for distances calculation of unrooted trees. The Robinson-Foulds distances (Robinson and Foulds, 1981) and percentage of branches similarity were calculated to indicate comparison results.

In silico phylogeny design for the PI guide tree

We used BEAST2 (Bouckaert et al., 2014) to estimate divergence time tree for PI calculation. The species divergence time tree was regenerated following the online tutorial Calibrated Species Trees on Taming the BEAST (Barido-Sottani et al., 2018). The gamma category count was set as 4, the substitution model was set as HKY, the mutation rate of strict clock was set as 0.001, and MCMC chain length was set as 500 million.

Calculation PI value and phylogenetic noise

PI value and phylogenetic noise of markers were calculated on PhyDesign (Lopez-Giraldez and Townsend, 2011), using HyPhy to calculate the rates (Pond et al., 2005). Nexus format files with partition blocks and the guide tree of the species divergence time tree were uploaded onto the webserver. There were four categories of phylogenetic markers were studied for PI values and phylogenetic noise: single-gene markers, concatenated-11-genes markers, concatenated-genes markers from the shortest

one (min2max), and concatenated-genes markers from the largest one (max2min), as listed in the Table 2.

Results and discussion

Reconstruction of phylogenetic relationships of nematodes

We reconstructed the phylogenetic trees of 93 nematodes species (Table 4.1) including two recently sequenced lance nematode species (*Hoplolaimus columbus* and *H. galeatus*) as references for further study. The reference trees were generated with 12 protein-coding genes. The concatenated-genes were analyzed by both Maximum Likelihood (ML) and Bayesian Inference (BI) methods, as shown in Figure 4.1 and Figure 4.2. Both trees have strong support values for most branches. We observed two methods agree with each other mostly. Only one difference was found between two results: in the BI tree, *Aelurostronglus abstrusus* was grouped with *Parafilaroides normani*; but in the ML tree, *A. abstrusus* was in the same clade of *Protostongylus rufescens*. This difference may be caused by long-branch attraction. Those reference trees also have similar phylogenetic results as published results (Park et al., 2011; Sultana et al., 2013; Humphreys-Pereira and Elling, 2014; Phillips et al., 2016; Palomares-Rius et al., 2017).

Phylogenetic tree comparison among different markers

We conducted similarity analyses of phylogenetic results using Excel to present the differences (Figure 4.4). Alignment lengths of each markers (Table 4.2) were presented in the figure as well. All ML trees were referred to the ML 12-genes tree and BI trees were compared with the BI 12-genes tree. For the same marker, both ML and BI present similar similarity value. When using single-gene markers to conduct phylogenetic analysis, the shortest nad4l gene shows the lowest similarity to the reference tree. The nad6 gene though has larger lengths than the atp6 gene, and its phylogenetic similarity to the reference is over 5% smaller than that of the atp6 gene. On the other side, markers of the nad1, nad4, and nad5 genes, which have the longest alignment length, have the highest similarities among 12 PCG markers. The traditional marker, the cox1 and cytb genes do not present the best similarities among those markers, which indicate that they may be not best representing the phylogenetic signal of mitochondrial genome. The single-gene markers also present a range of diverse phylogenetic similarities referring to the 12-gene tree, which may indicate each gene has different mutation rates even though they are located within the same chromosome.

We also tested concatenated-genes markers of different patterns. One is to concatenate 11 of 12 genes, and to generate 12 markers, each of them missing one gene. All 11-genes tree are over 95% similar as the 12-genes tree. However, the results show that when the cox2 gene or the nad3 gene is missing, the phylogenetic results are still the same as the 12-genes tree. The other two are to concatenate genes from single gene

incrementally to 11-genes, either starting from the largest gene marker or from the smallest one. Generally, when both the alignment length and the number of genes increases, the similarity is also increasing, except two cases. When adding gene markers from the smallest (the nad4l gene) to the fifth smallest one (the cytb gene), the similarity of this 5-genes marker in both the ML and BI methods are smaller than the 4-genes marker next to it. This decreasing is also found between after and before adding the cox1 gene marker. If the concatenated-genes were started from the longest gene, nad5, the similarity fluctuation is even and gentle among the following several markers.

In a summary of the results from phylogenetic tree comparison, the phylogenetic similarities reveal that genes may contain different phylogenetic signals. Although different sequence lengths of phylogenetic markers may influence the phylogenetic signal, it doesn't mean a longer sequence is sufficient to give better phylogenetic results.

Species divergence time tree estimation

We conducted a species divergence time tree (DT tree) with a strict molecular clock model as in Figure 4.3. Some differences between the DT tree and the BI tree were found. First of all, *Acanthocheilonema viteae* is located differently. Secondly, in the DT tree, *Ancylostoma* spp. were grouped with *Oesophagostomum* spp., while in the BI tree, they were distanced. The third one is about the relationships between *Ascaris* spp. with *Parascaris univalens*. Moreover, *Steinernema carpocapsae* was located very differently

in the DT tree from that of the BI tree. Although some differences were found, the tree generally agrees with the BI and ML trees.

Phylogenetic informativeness (PI) and noise of mitochondrial markers

Both phylogenetic informativeness and noise were calculated on PhyDesign webserver and guided by the species divergence tree. In Figure 4.5, the PI values of markers gradually change along the time axis and all markers ranked vertically based on their maximum PI value in the time range between 4 million years ago and 10 million years ago. The high spike close to time 0, as described on the PhyDesign webserver, is distanced from the peak of the informativeness curve in most markers. The results in Figure 4.5A show the nad5 and nad4 genes have higher informativeness than the traditional markers of the cox1 gene and the cytb gene. The atp6 gene presented a dramatically different mutation pattern than other genes, which may contain unique phylogenetic signals. The nad3 and nad4l contain the lowest phylogenetic informativeness, which agrees with the afore result of phylogenetic tree comparison. In Figure 4.5B, the 11-genes markers all present high PI value. However, if the marker is missing nad5 (nonad5), the PI value drops dramatically. It also proves that nad5 contains relatively higher phylogenetic signals. In Figure 4.5C and 4.5D, two different patterns of concatenated-genes indicate that the concatenated-genes marker has higher PI value if additional genes are added. These results present a different character of phylogenetic signals from the tree comparison results, since PI value is gradually accumulated along

with the number of genes used while topological similarity may have fluctuation during increasing number of used genes.

In Figure 4.6, phylogenetic noises of markers are plotted. Single-gene markers represent different noise and correct probability: the nad4l and nad3 genes are considered the most unreliable markers among 12 PCGs, which have the highest probability of noise and polytomy and lowest correct probability. Two traditional markers, the cytb and cox1 gene, are the two markers with both high correct probability and low noise. However, their polytomy probabilities may indicate that they have less resolution for phylogenetic analysis than nad4 and nad5 markers. The nad4 and nad5 genes could be considered as two good phylogenetic markers since they have high correct probability as well as relatively low noise and polytomy probability. The nad1 gene marker, which presented relatively high similarity in tree comparison method, show low correct probability but high polytomy probability. Concatenated-gene markers show that more genes concatenated in a marker indicate a higher correct probability. All 11-genes markers present a stable probability on high correct and low noise signal. The probability of polytomy is suppressed as well.

In summary, the PI value and phylogenetic signal noise reveal that nad4 and nad5 could be two candidates for phylogenetic analysis. Traditional markers such as cox1 or cytb may not be the optimum options for phylogenetic analyses. In fact, there has been

reported that on nematode population study the molecular diversity on cox1 or cytb genes is lower than nad4 gene (Cabasan et al., 2018).

Conclusion

In this investigation, we studied the phylogenetic validation of mitochondrial protein-coding genes using two methods of tree comparison and phylogenetic informativeness. Results from two methods agree with each other on several aspects: (1) Longer sequence length is not sufficient to ensure a better phylogenetic analyses; (2) Concatenated-genes markers with suitable genes offer better phylogeny than each of the single genes; (3) Nad5 and nad4 genes could be relatively better phylogenetic markers of nematodes; (4) Traditional mitochondrial markers, such as cox1 or cytb gene are relatively more stable since both of them have the smallest phylogenetic noise. However, they may not offer adequate resolution for some phylogenetic analyses, such as population genetics or biogeographic study; (5) We do not suggest cox2, nad3, nad4l, or nad6 gene to be an independent phylogenetic marker for nematodes.

References

Andersson SGE, Karlberg O, Canbäck B, Kurland CG. 2003. On the origin of mitochondria: a genomics perspective. Philos Trans R Soc Lond B Biol Sci. 358: 165–179.

Armstrong MR, Blok VC, Phillips MS. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. Genetics. 154:181-192.

Barido-Sottani J et al. 2018. Taming the BEAST – A community teaching material resource for BEAST 2. Systematic Biology, 67(1), 170-174.

Beck-Azevedo JL, Hyman BC. 1993. Molecular characterization of lengthy mitochondrial DNA duplications from the parasitic nematode *Romanomermis culicivorax*. Genetics. 133: 933–942.

Bernt A, et al. 2013. MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation. Molecular Phylogenetics and Evolution. 69:313-319.

Blaxter ML, et al. 1998. A molecular evolutionary framework for the phylum Nematoda. Nature. 392: 71-75.

Bongers T, Ferris H. 1999. Nematode community structure as a bioindicator in environmental monitoring. Trends Ecol Evol. 14(6):224-228.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ.2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10(4):e1003537.

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG. 2004. Computational inference of scenarios for alpha-proteobacterial genome evolution. Proc Natl Acad Sci USA. 101:9722-9727

Cabasan MTN, Kumar A., and Waele DD. 2018. Effects of initial nematode population density and water regime on resistance and tolerance to the rice root-knot nematode *Meloidogyne graminicola* in African and Asian rice genotypes. International Journal of Pest Management. 64, 3:252-261.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 99: 45-56.

Chitwood, B. G. and M. B. Chitwood. 1933. Nematodes parasitic in Philippine cockroaches. Philippine Journal of Science 52:381–393.

Chomyn A, Attardi G. 1987. Mitochondrial gene products. Curr Topic Bioenergetics. 15:295-329.

Cock PA, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25:1422-1423.

De Ley P, Blaxter ML. 2002. Systematic position and phylogeny. In: The Biology of Nematodes, D.L. Lee, ed., London: Taylor and Francis. p.1–30.

Dereeper A, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:465-469.

Dereeper A, Audic S, Claverie JM, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol. 10:8.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.

Dominik R, Heitlinger EG, Taraschewski H, Nadler SA, Blaxter ML.2012. The phylogenetics of Anguillicolidae (Nematoda: Anguillicoloidea), swim bladder parasites of eels. BMC Evolutionary Biology, 12:60.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology. 59:307-321.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Res. 41: e129.

Holterman M, et al. 2006. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. Mol Biol Evol. 23:1792-800.

Howe DK, Denver DR. 2008. Muller's Ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. BMC Evol Biol.8:62.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17(8):754–755,

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. Molecular Biology and Evolution, 33(6), 1635-1638.

Humphreys-Pereira DA, Elling AA. 2014. Mitochondrial genomes of *Meloidogyne chitwoodi* and *M. incognita* (Nematoda: Tylenchina): comparative analysis,

gene order and phylogenetic relationships with other nematodes. Mol Biochem Parasitol.194: 20-32.

Kim J, et al. 2017. Phylogenetic analysis of two Plectus mitochondrial genomes (Nematoda: Plectida) supports a sister group relationship between Plectida and Rhabditida within Chromadorea. Mol Phylogenet Evol. 107:90-102.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41:22-28.

Lagisz M, Poulin R., Nakagawa S. 2013. You are where you live: Parasitic nematode mitochondrial genome size is associated with the thermal environment generated by hosts. Journal of Evolutionary Biology 26(3).

Lavrov DV, Brown WM. 2001. *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. Genetics. 157: 621-637.

Lefort V, Longueville J, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution. 34:2422-2424.

Liu GH, et al. 2012. Characterization of the complete mitochondrial genomes of two whipworms Trichuris ovis and Trichuris discolor (Nematoda: Trichuridae). Infect. Genet. Evol. 12: 1635-1641.

Lopez-Giraldez F, Townsend JP. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol. Biol., 11: 152.

Ma X, Agudelo P, Mulluer JD, Knap HT. 2011. Molecular Characterization and Phylogenetic Analysis of *Hoplolaimus stephanus*. J Nematol. 43:25-34.

van Megen H, et al. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology. 11:927-950.

Palomares-Rius JE, Cantalapiedra-Navarrete C, Archidona-Yuste A, Blok VC, Castillo P. 2017. Mitochondrial genome diversity in dagger and needle nematodes (Nematoda: Longidoridae). Sci Rep. 7: 41813.

Park JK et al. 2011. Monophyly of clade III nematodes is not supported by phylogenetic analysis of complete mitochondrial genome sequences. BMC Genomics. 12:392.

Phillips WS et al. 2016. The mitochondrial genome of *Globodera ellingtonae* is composed of two circles with segregated gene content and differential copy numbers. BMC Genomics.17:706.

Platt, H.M. 1994. Foreward. In: The phylogenetic systematics of free-living nematodes, S. Lorenzen, ed. (London: The Ray Society).

Pond, SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics, 21(5): 676-679.

Powers TO, Todd TC, Burnell AM, Murray PCB., Fleming C., Szalanski AL, Adams B A, Harris TS. 1997. The rDNA Internal Transcribed Spacer Region as a Taxonomic Marker for Nematodes. J Nematol. Dec; 29,4: 441–450.

Raymond MR, Wharton DA. 2016. The ability to survive intracellular freezing in nematodes is related to the pattern and distribution of ice formed. Journal of Experimental Biology. jeb.137190.

Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Ronquist, F., Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B., Liu L, Suchard MA., Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology, 61(3): 539-42.

Sagan L. 1967. On the origin of mitosing cells. J Theor Biol. 14:225-247.

Saint André A et al. 2002. The role of endosymbiotic *Wolbachia bacteria* in the pathogenesis of river blindness. Science. 295(5561):1892-1895.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Annals of the Entomological Society of America, 87, 651–701.

Stajich JE, et al. 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12: 1611-1618.

Suga K, Mark Welch DB, Tanaka Y, Sakakura Y, Hagiwara A. 2008. Two circular chromosomes of unequal copy number make up the mitochondrial genome of the rotifer *Brachionus plicatilis*. Mol Biol Evol 25: 1129-1137.

Sultana T, et al. 2013. Comparative analysis of complete mitochondrial genome sequences confirms independent origins of plant-parasitic nematodes. BMC Evol Biol. 13: 12.

Townsend JP. 2007. Profiling phylogenetic informativeness. Syst Biol. 2007 Apr;56(2):222-31.

Townsend JP, Su Z, Tekle YI. 2012. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. Syst Biol. 61(5):835-849.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi‐gene datasets with character set and codon information. Cladistics. 27:171-180.

Zou H, Jakovlić I, Chen R, Zhang D, Zhang J, Li W-Xand Wang GT. The complete mitochondrial genome of parasitic nematode *Camallanus cotti*: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class. BMC Genomics. 2017;18: 840.

Tables and Figures

Table 4.1. Nematode species used in the phylogenetic informativeness investigation and their GenBank accession numbers.

Species	Accessions	Sizes (bp)	Species	Accessions	Sizes (bp)
Acanthocheilonema viteae	NC_016197	13,724	*Meloidogyne javanica*	NC_026556.1	18,291
Aelurostrongylus abstrusus	NC_019571.1	13,913	*Nippostrongylus brasiliensis*	NC_033886.1	13,355
Ancylostoma ceylanicum	NC_035142.1	13,660	*Oesophagostomum dentatum*	GQ888716.1	13,869
Ancylostoma tubaeforme	NC_034289.1	13,730	*Oesophagostomum quadrispinulatum*	NC_014181.1	13,681
Agamermis.spBH2006	NC_008231	16,561	*Onchocerca flexuosa*	NC_016172.1	13,672
Angiostrongylus cantonensis	GQ398121.1	13,497	*Onchocerca ochengi*	KX181290.2	13,744
Angiostrongylus costaricensis	NC_013067.1	13,585	*Oscheius chongmingensis*	KP257594.1	15,413
Angiostrongylus malaysiensis	NC_030332.1	13,516	*Parafilaroides normani*	NC_024656.1	13,414
Angiostrongylus vasorum	NC_018602.1	13,422	*Paralongidorus litoralis*	NC_033868.1	12,763
Anisakis pegreffii	LC222461.1	14,002	*Parascaris univalens*	KM216010.1	14,350
Aphelenchoides besseyi	KJ739799.1	16,216	*Passalurus ambiguus*	NC_028345.1	14,023
Ascaris lumbricoides	HQ704900.1	14,303	*Philometroides sanguineus*	NC_024931.1	14,378
Ascaris suum	HQ704901.1	14,311	*Pratylenchus vulnus*	NC_020434.1	21,656
Baylisascaris procyonis	NC_016200.1	14,781	*Pristionchus pacificus*	JF414117.1	15,954
Brugia malayi	NC_004298.1	13,657	*Protostrongylus rufescens*	NC_023262.1	13,619
Bursaphelenchus xylophilus	JQ514068.1	12,945	*Pseudoterranova azarasi*	NC_027163.1	13,954
Caenorhabditis briggsae	AC186293.1	14,420	*Radopholus similis*	FN313571.1	16,791
Caenorhabditis elegans	NC_001328.1	13,794	*Rhigonema thysanophora*	NC_024020.1	15,015
Caenorhabditis nigoni	KP259621.2	13,856	*Romanomermis culicivorax*	NC_008640.1	26,194
Caenorhabditis remanei	KR709159.1	13,977	*Romanomermis iyengari*	EF175764.1	18,919
Caenorhabditis sinica	EU407780.1	13,537	*Romanomermis nielseni*	EF175763	15,546
Caenorhabditis tropicalis	NC_025756.1	13,874	*Rotylenchulus reniformis*	CM003310.1	24,572
Contracaecum rudolphii	FJ905109.1	14,022	*Setaria digitata*	KY284626.1	13,814
Cooperia oncophora	NC_004806.1	13,636	*Spirocerca lupi*	NC_021135.1	13,780
Cucullanus robustus	GQ332426.1	13,972	*Steinernema carpocapsae*	NC_005941.1	13,925
Cylicocyclus nassatus	NC_032299.1	13,846	*Strelkovimermis spiculatus*	NC_008047.1	18,030
Dirofilaria immitis	NC_005305.1	13,814	*Strongyloides papillosus*	LC050210.1	13,909
Enterobius vermicularis	EU281143.1	14,010	*Strongyloides ratti*	NC_028623.1	16,609
Globodera ellingtonae (I)	KU726971.1	17,757	*Strongyloides stercoralis*	NC_028624.1	13,751
Globodera ellingtonae (II)	KU726972.1	14,365	*Strongyloides venezuelensis*	LC050213.1	15,956
Gnathostoma spinigerum	NC_027726.1	14,079	*Strongylus equinus*	NC_026868.1	14,545
Gongylonema pulchrum	NC_026687.1	13,798	*Thelazia callipaeda*	NC_018363.1	13,668
Haemonchus contortus	NC_010383.2	14,055	*Toxascaris leonina*	NC_023504.1	14,310
Heliconema longissimum	NC_016127.1	13,610	*Toxocara canis*	NC_010690.1	14,322
Heterodera glycines	HM640930.1	14,915	*Toxocara cati*	NC_010773.1	14,029
Heterorhabditis bacteriophora	NC_008534.1	18,128	*Toxocara malaysiensis*	NC_010527.1	14,266
Hexamermis agrotis	EF368011.1	24,606	*Trichinella spiralis*	KM357422.1	16,584
Hoplolaimus columbus	MH657221	25,228	*Trichostrongylus axei*	NC_013824.1	13,653
Hoplolaimus galeatus	MK119781	25,218	*Trichuris sp.GHL2013*	KC461179.1	14,147
Koerneria sudhausi	NC_029233.1	16,005	*Trichuris suis*	NC_017747.1	14,436
Loa loa	HQ186250.1	13,590	*Trichuris trichiura*	NC_017750.1	14,046
Longidorus vineacola	NC_033867.1	13,519	*Triodontophorus brevicauda*	NC_026729.1	14,305
Meloidogyne arenaria	NC_026554.1	17,580	*Wellcomia siamensis*	NC_016129.1	14,128
Meloidogyne chitwood	KJ476150.1	18,201	*Wuchereria bancrofti*	HQ184469.1	13,636
Meloidogyne enterolobii	KP202351.1	17,053	*Xiphinema americanum*	AY382608.1	12,626
Meloidogyne graminicola	KJ139963.1	19,589	*Xiphinema pachtaicum*	NC_033870.1	12,489
Meloidogyne incognita	KJ476151.1	17,662	*Xiphinema rivesi*	NC_033869.1	12,624

Table 4.2. Alignment lengths of phylogenetic markers and their concatenation patterns used in this study

marker	alignment length (bp)	marker	alignment length (bp)
nad4l	498	nonad5	14505
nad3	673	nonad4	15468
atp6	1091	nonad1	15486
nad6	1172	nocox1	15515
ctyb	1255	nonad2	15642
cox3	1424	nocox2	15732
cox2	1515	nocox3	15823
nad2	1605	nocytb	15992
cox1	1732	nonad6	16075
nad1	1761	noatp6	16156
nad4	1779	nonad3	16574
nad5	2742	nonad4l	16749
nad4l=>2	1171	nad4l+nad3	
nad4l=>3	2262	nad4l+nad3+atp6	
nad4l=>4	3434	nad4l+nad3+atp6+nad6	
nad4l=>5	4689	nad4l+nad3+atp6+nad6+ctyb	
nad4l=>6	6113	nad4l+nad3+atp6+nad6+ctyb+cox3	
nad4l=>7	7628	nad4l+nad3+atp6+nad6+ctyb+cox3+cox2	
nad4l=>8	9233	nad4l+nad3+atp6+nad6+ctyb+cox3+cox2+nad2	
nad4l=>9	10965	nad4l+nad3+atp6+nad6+ctyb+cox3+cox2+nad2+cox1	
nad4l=>10	12726	nad4l+nad3+atp6+nad6+ctyb+cox3+cox2+nad2+cox1+nad1	
nad4l=>11	14505	nad4l+nad3+atp6+nad6+ctyb+cox3+cox2+nad2+cox1+nad1+nad4	
nad5=>2	4521	nad5+nad4	
nad5=>3	6282	nad5+nad4+nad1	
nad5=>4	8014	nad5+nad4+nad1+cox1	
nad5=>5	9619	nad5+nad4+nad1+cox1+nad2	
nad5=>6	11134	nad5+nad4+nad1+cox1+nad2+cox2	
nad5=>7	12558	nad5+nad4+nad1+cox1+nad2+cox2+cox3	
nad5=>8	13813	nad5+nad4+nad1+cox1+nad2+cox2+cox3+ctyb	
nad5=>9	14985	nad5+nad4+nad1+cox1+nad2+cox2+cox3+ctyb+nad6	
nad5=>10	16076	nad5+nad4+nad1+cox1+nad2+cox2+cox3+ctyb+nad6+atp6	
nad5=>11	16749	nad5+nad4+nad1+cox1+nad2+cox2+cox3+ctyb+nad6+atp6+nad3	
12gene	17247	atp6+cox1+cox2+cox3+cytb+nad1+nad2+nad3+nad4+nad4l+nad5+nad6	

Table 4.3. Topological similarity of phylogenetic results using Maximum Likelihood method

Maximum Likelihood							
source	ref	E.size	nRF	RF	maxRF	src-br+	ref-br+
12gene.ml	12gene.ml	93	0	0	180	1	1
atp6	12gene.ml	93	0.34	62	180	0.83	0.83
cox1	12gene.ml	93	0.38	68	180	0.81	0.81
cox2	12gene.ml	93	0.43	78	180	0.79	0.79
cox3	12gene.ml	93	0.34	62	180	0.83	0.83
ctyb	12gene.ml	93	0.37	66	180	0.82	0.82
nad1	12gene.ml	93	0.22	40	180	0.89	0.89
nad2	12gene.ml	93	0.47	84	180	0.77	0.77
nad3	12gene.ml	93	0.42	76	180	0.79	0.79
nad4	12gene.ml	93	0.27	48	180	0.87	0.87
nad4l	12gene.ml	93	0.59	106	180	0.71	0.71
nad5	12gene.ml	93	0.28	50	180	0.86	0.86
nad6	12gene.ml	93	0.46	82	180	0.77	0.77
nad4l=>2gene	12gene.ml	93	0.36	64	180	0.82	0.82
nad4l=>3genes	12gene.ml	93	0.32	58	180	0.84	0.84
nad4l=>4genes	12gene.ml	93	0.22	40	180	0.89	0.89
nad4l=>5genes	12gene.ml	93	0.27	48	180	0.87	0.87
nad4l=>6genes	12gene.ml	93	0.14	26	180	0.93	0.93
nad4l=>7genes	12gene.ml	93	0.16	28	180	0.92	0.92
nad4l=>8genes	12gene.ml	93	0.12	22	180	0.94	0.94
nad4l=>9genes	12gene.ml	93	0.14	26	180	0.93	0.93
nad4l=>10genes	12gene.ml	93	0.11	20	180	0.95	0.95
nad5=>2genes	12gene.ml	93	0.22	40	180	0.89	0.89
nad5=>3genes	12gene.ml	93	0.17	30	180	0.92	0.92
nad5=>4genes	12gene.ml	93	0.14	26	180	0.93	0.93
nad5=>5genes	12gene.ml	93	0.09	16	180	0.96	0.96
nad5=>6genes	12gene.ml	93	0.1	18	180	0.95	0.95
nad5=>7genes	12gene.ml	93	0.09	16	180	0.96	0.96
nad5=>8genes	12gene.ml	93	0.03	6	180	0.98	0.98
nad5=>9genes	12gene.ml	93	0.01	2	180	0.99	0.99
nad5=>10genes	12gene.ml	93	0.01	2	180	0.99	0.99
noatp6	12gene.ml	93	0.04	8	180	0.98	0.98
nocox1	12gene.ml	93	0.04	8	180	0.98	0.98
nocox2	12gene.ml	93	0.01	2	180	0.99	0.99
nocox3	12gene.ml	93	0.03	6	180	0.98	0.98
nocytb	12gene.ml	93	0.07	12	180	0.97	0.97
nonad1	12gene.ml	93	0.08	14	180	0.96	0.96
nonad2	12gene.ml	93	0.07	12	180	0.97	0.97
nonad3	12gene.ml	93	0	0	180	1	1
nonad4	12gene.ml	93	0.09	16	180	0.96	0.96
nonad4l	12gene.ml	93	0.04	8	180	0.98	0.98
nonad5	12gene.ml	93	0.06	10	180	0.97	0.97
nonad6	12gene.ml	93	0.02	4	180	0.99	0.99

Table 4.4. Topological similarity of phylogenetic results using Bayesian Inference method

Bayesian Inference							
source	ref	E.size	nRF	RF	maxRF	src-br+	ref-br+
12gene.bi	12gene.bi	93	0	0	180	1	1
atp6.bi	12gene.bi	93	0.35	59	171	0.86	0.81
cox1.bi	12gene.bi	93	0.36	64	178	0.83	0.82
cox2.bi	12gene.bi	93	0.41	72	176	0.81	0.79
cox3.bi	12gene.bi	93	0.31	53	169	0.88	0.82
ctyb.bi	12gene.bi	93	0.34	59	173	0.85	0.82
nad1.bi	12gene.bi	93	0.26	46	176	0.88	0.86
nad2.bi	12gene.bi	93	0.41	71	173	0.82	0.79
nad3.bi	12gene.bi	93	0.43	66	154	0.87	0.75
nad4.bi	12gene.bi	93	0.27	47	177	0.88	0.86
nad4l.bi	12gene.bi	93	0.59	85	145	0.83	0.67
nad5.bi	12gene.bi	93	0.32	57	177	0.85	0.84
nad6.bi	12gene.bi	93	0.45	67	149	0.88	0.73
nad4l=>2.bi	12gene.bi	93	0.41	66	162	0.85	0.77
nad4l=>3.bi	12gene.bi	93	0.3	54	178	0.86	0.85
nad4l=>4.bi	12gene.bi	93	0.24	42	176	0.89	0.87
nad4l=>5.bi	12gene.bi	93	0.25	45	177	0.88	0.87
nad4l=>6.bi	12gene.bi	93	0.13	24	180	0.93	0.93
nad4l=>7.bi	12gene.bi	93	0.13	24	180	0.93	0.93
nad4l=>8.bi	12gene.bi	93	0.09	16	180	0.96	0.96
nad4l=>9.bi	12gene.bi	93	0.15	27	179	0.93	0.92
nad4l=>10.bi	12gene.bi	93	0.1	18	180	0.95	0.95
nad5=>2.bi	12gene.bi	93	0.23	41	177	0.89	0.88
nad5=>3.bi	12gene.bi	93	0.16	28	180	0.92	0.92
nad5=>4.bi	12gene.bi	93	0.13	23	179	0.94	0.93
nad5=>5.bi	12gene.bi	93	0.09	16	180	0.96	0.96
nad5=>6.bi	12gene.bi	93	0.09	16	180	0.96	0.96
nad5=>7.bi	12gene.bi	93	0.05	9	179	0.98	0.97
nad5=>8.bi	12gene.bi	93	0.01	2	180	0.99	0.99
nad5=>9.bi	12gene.bi	93	0.01	2	180	0.99	0.99
nad5=>10.bi	12gene.bi	93	0.06	11	179	0.97	0.97
noatp6.bi	12gene.bi	93	0.01	2	180	0.99	0.99
nocox1.bi	12gene.bi	93	0.04	8	180	0.98	0.98
nocox2.bi	12gene.bi	93	0	0	180	1	1
nocox3.bi	12gene.bi	93	0.03	6	180	0.98	0.98
nocytb.bi	12gene.bi	93	0.07	12	180	0.97	0.97
nonad1.bi	12gene.bi	93	0.08	15	179	0.96	0.96
nonad2.bi	12gene.bi	93	0.02	4	180	0.99	0.99
nonad3.bi	12gene.bi	93	0	0	180	1	1
nonad4l.bi	12gene.bi	93	0.01	2	180	0.99	0.99
nonad4.bi	12gene.bi	93	0.03	5	179	0.99	0.98
nonad5.bi	12gene.bi	93	0.09	16	180	0.96	0.96
nonad6.bi	12gene.bi	93	0.02	3	179	0.99	0.99

Table 4.5. Phylogenetic noise of all markers in this study

	t=0.1,T=2			t=0.01,T=4		
Loci	Prob Correct	Prob Noise	Prob Polytomy	Prob Correct	Prob Noise	Prob Polytomy
nad4l	0.33516	0.38119	0.28365	0.23919	0.54636	0.21445
nad3	0.36304	0.39656	0.2404	0.25377	0.56359	0.18264
atp6	0.48629	0.39544	0.11827	0.29809	0.61121	0.0907
nad6	0.42585	0.39956	0.17459	0.27772	0.58903	0.13325
cytb	0.53391	0.29879	0.1673	0.28695	0.58736	0.12569
cox3	0.47335	0.33438	0.19227	0.27639	0.58086	0.14275
cox2	0.44588	0.36428	0.18984	0.27488	0.58095	0.14417
nad2	0.4963	0.38252	0.12118	0.29822	0.60662	0.09516
cox1	0.58366	0.2673	0.14904	0.29368	0.58901	0.11731
nad1	0.49386	0.33577	0.17037	0.28409	0.58818	0.12773
nad4	0.54672	0.32598	0.1273	0.29916	0.60093	0.09991
nad5	0.56383	0.34162	0.09455	0.31025	0.61217	0.07758
nad5=>2	0.67487	0.25716	0.06797	0.32397	0.6143	0.06173
nad5=>3	0.73238	0.21154	0.05608	0.33037	0.61371	0.05592
nad5=>4	0.80293	0.1544	0.04267	0.33809	0.61108	0.05083
nad5=>5	0.8311	0.13471	0.03419	0.34306	0.6118	0.04514
nad5=>6	0.85133	0.11864	0.03003	0.34596	0.61079	0.04325
nad5=>7	0.87273	0.10136	0.02591	0.34891	0.60952	0.04157
nad5=>8	0.89803	0.080907	0.021063	0.35275	0.60759	0.03966
nad5=>9	0.90706	0.074044	0.018896	0.35458	0.60726	0.03816
nad5=>10	0.91746	0.066505	0.016035	0.35674	0.6079	0.03536
nad5=>11	0.92233	0.062617	0.015053	0.35778	0.60743	0.03479
nad4l	0.33516	0.38119	0.28365	0.23919	0.54636	0.21445
nad4l=>2	0.44004	0.37528	0.18468	0.27587	0.58391	0.14022
nad4l=>3	0.56536	0.33835	0.09629	0.30908	0.61431	0.07661
nad4l=>4	0.61904	0.30081	0.08015	0.31664	0.61658	0.06678
nad4l=>5	0.70335	0.23256	0.06409	0.32534	0.61531	0.05935
nad4l=>6	0.75011	0.19562	0.05427	0.33048	0.61442	0.0551
nad4l=>7	0.78198	0.17114	0.04688	0.33461	0.61367	0.05172
nad4l=>8	0.81406	0.14863	0.03731	0.34005	0.61422	0.04573
nad4l=>9	0.86091	0.11034	0.02875	0.34591	0.61121	0.04288
nad4l=>10	0.88338	0.092462	0.024158	0.34935	0.60982	0.04083
nad4l=>11	0.90783	0.073311	0.018859	0.35383	0.60814	0.03803
nonad5	0.90783	0.073311	0.018859	0.35383	0.60814	0.03803
nonad4	0.90717	0.074773	0.018057	0.35475	0.60884	0.03641
nonad1	0.91277	0.070423	0.016807	0.35597	0.60847	0.03556
nocox1	0.90314	0.078554	0.018306	0.35441	0.60981	0.03578
nonad2	0.91512	0.067867	0.017013	0.3558	0.60754	0.03666
nocox2	0.91682	0.067071	0.016109	0.35659	0.60812	0.03529
nocox3	0.91448	0.069079	0.016441	0.35637	0.60832	0.03531
nocytb	0.90889	0.073735	0.017375	0.35542	0.60899	0.03559
nonad6	0.91914	0.065029	0.015831	0.35702	0.6075	0.03548
noatp6	0.91709	0.066142	0.016768	0.35661	0.60645	0.03694
nonad3	0.92167	0.063111	0.015219	0.3576	0.60746	0.03494
nonad4l	0.92233	0.062617	0.015053	0.35778	0.60743	0.03479
12gene	0.92628	0.059431	0.014289	0.35862	0.60699	0.03439

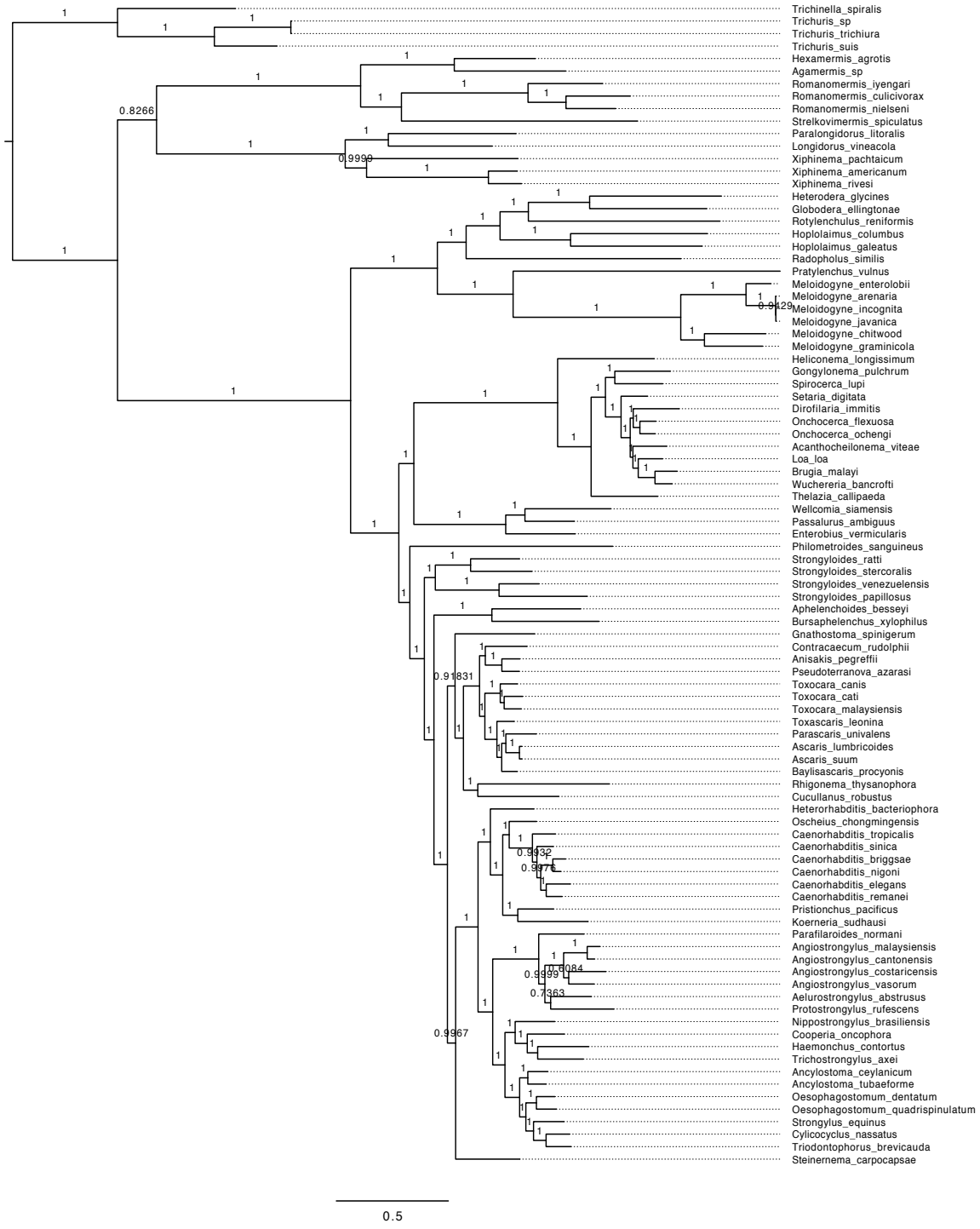

Figure 4.7. Phylogenetic tree of 12-concatenated-genes using Maximum Likelihood method

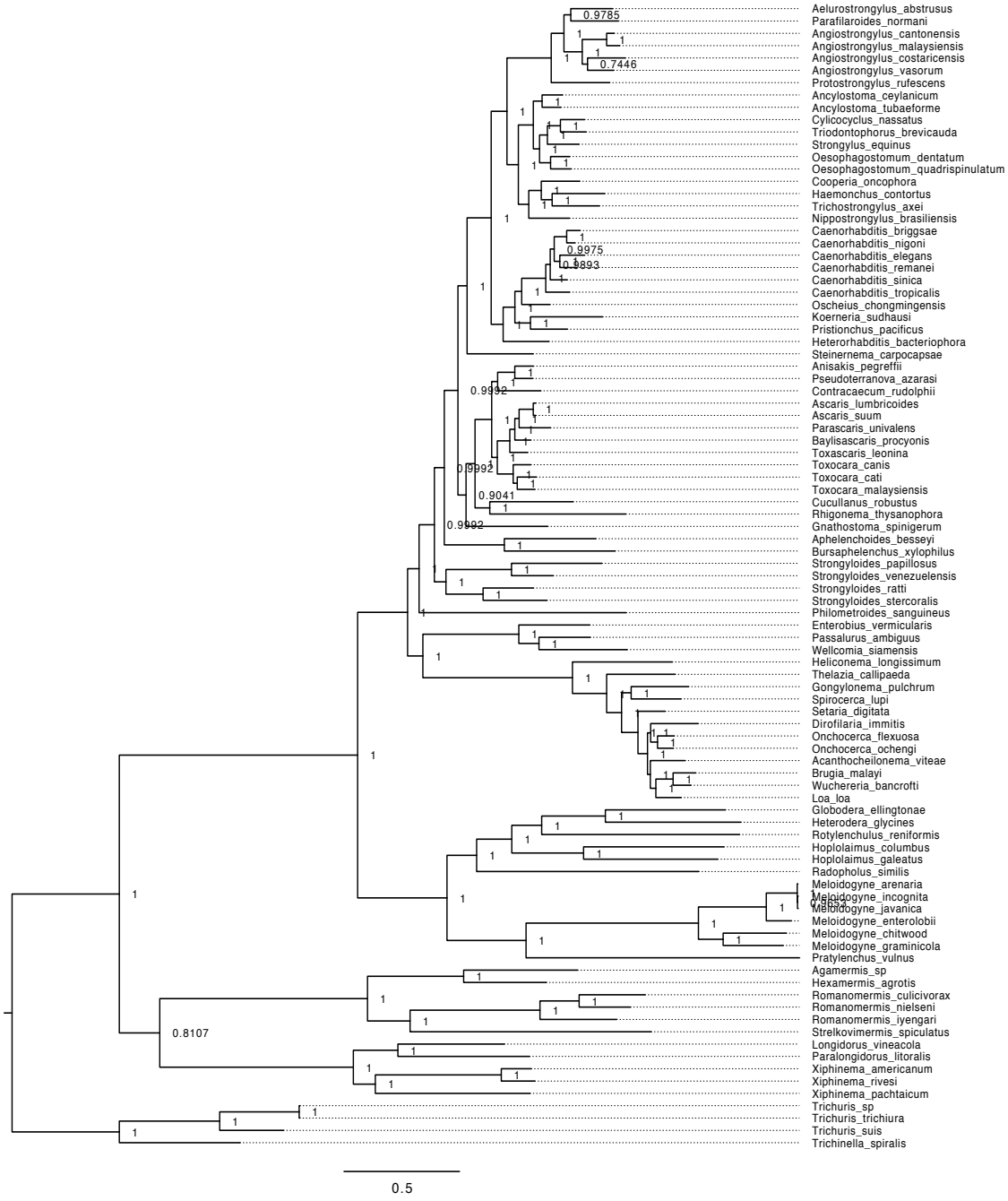

Figure 4.8. Phylogenetic tree of 12-concatenated-genes using Bayesian Inference method

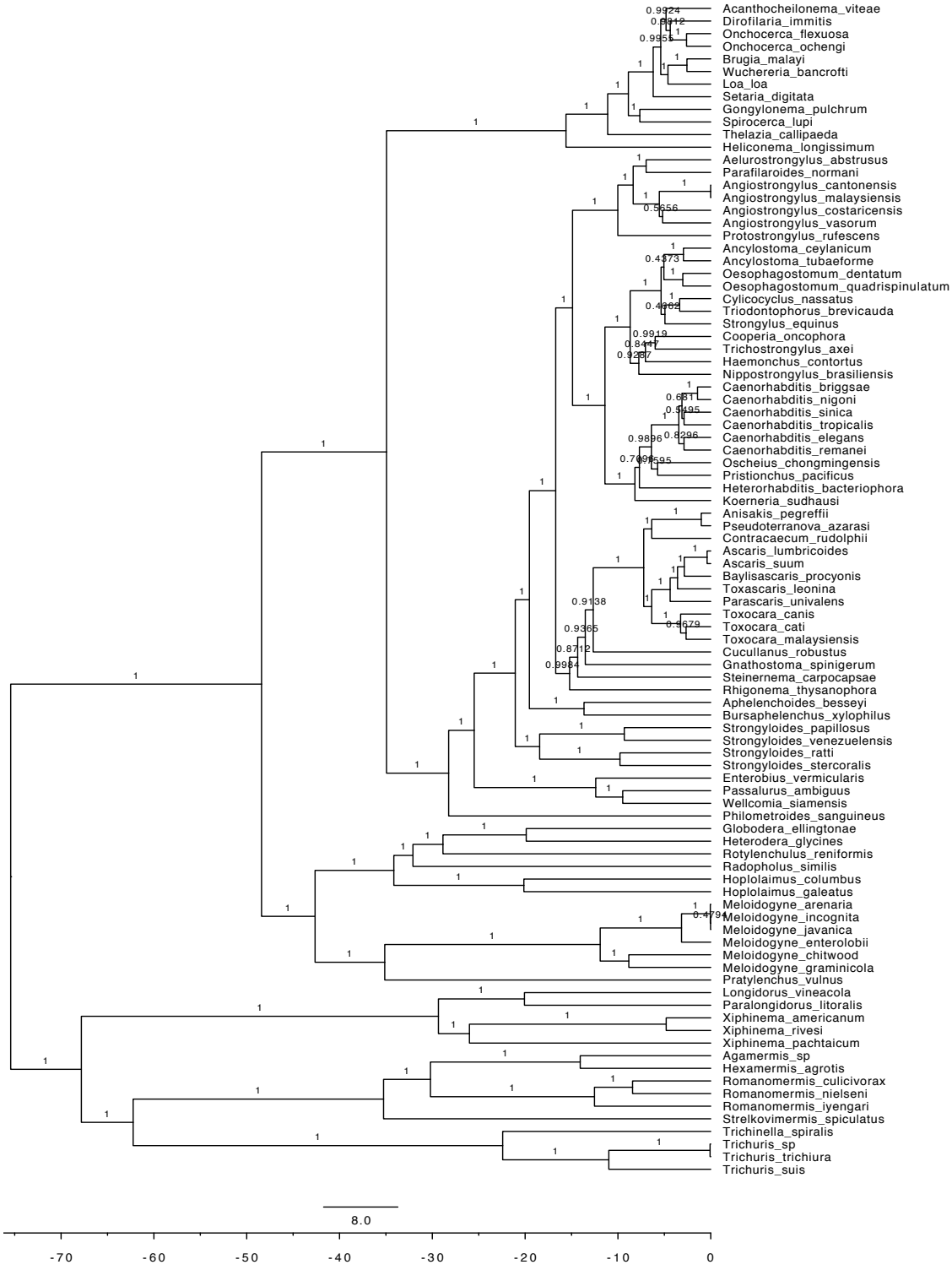

Figure 4.9. Species divergence time tree using strict molecular clock model 12 mitochondrial protein-coding genes

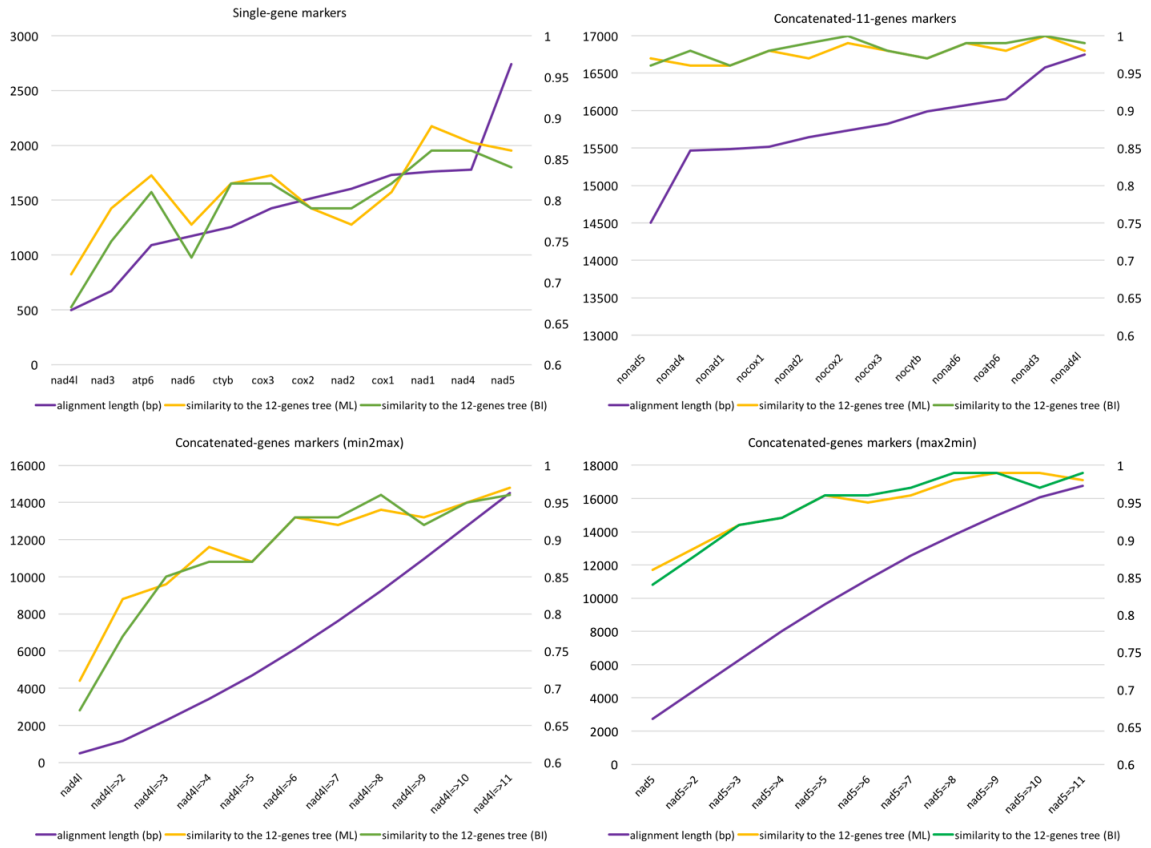

Figure 4.4. Topological comparison of phylogenetic markers referring to 12-genes tree using ETE3 method. Topological similarities were calculated by Robinson-Foulds Symmetric Distance and Branch Similarity.

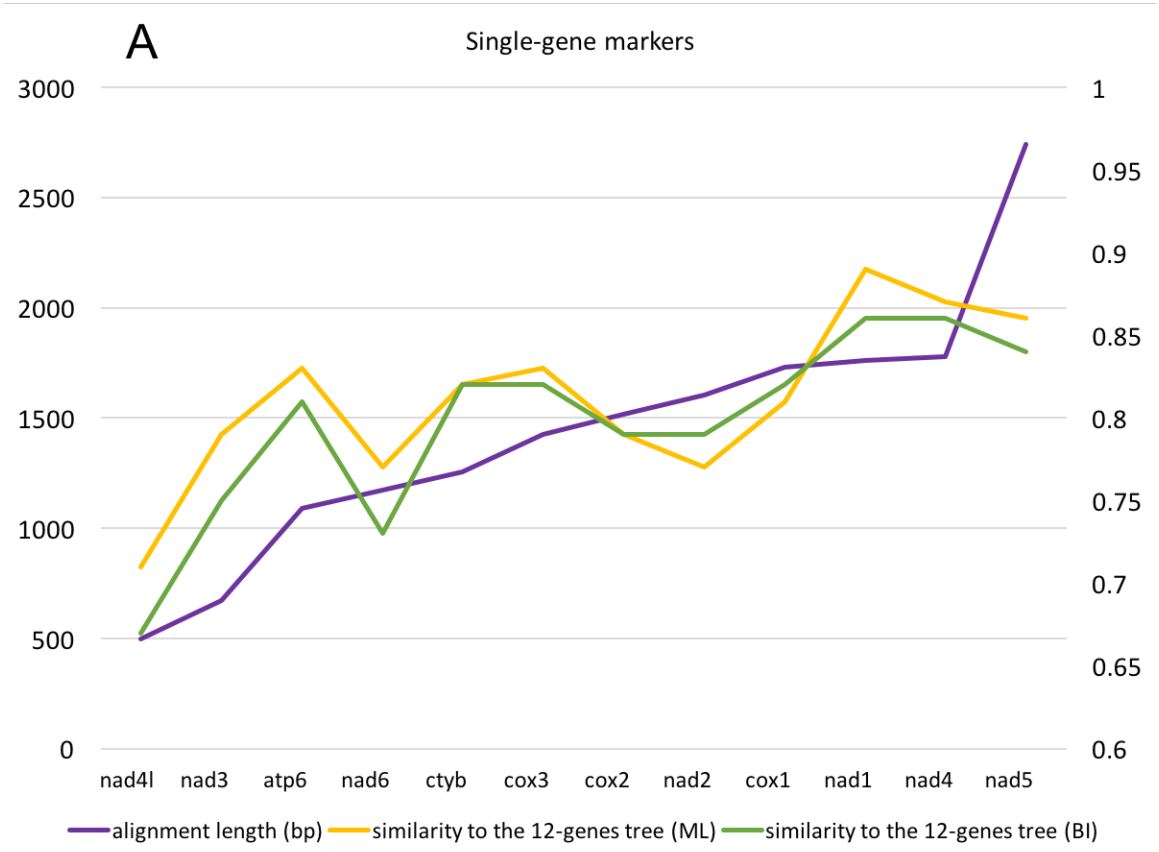

Figure 4.4A. Topological comparison of phylogenetic markers referring to 12-genes tree

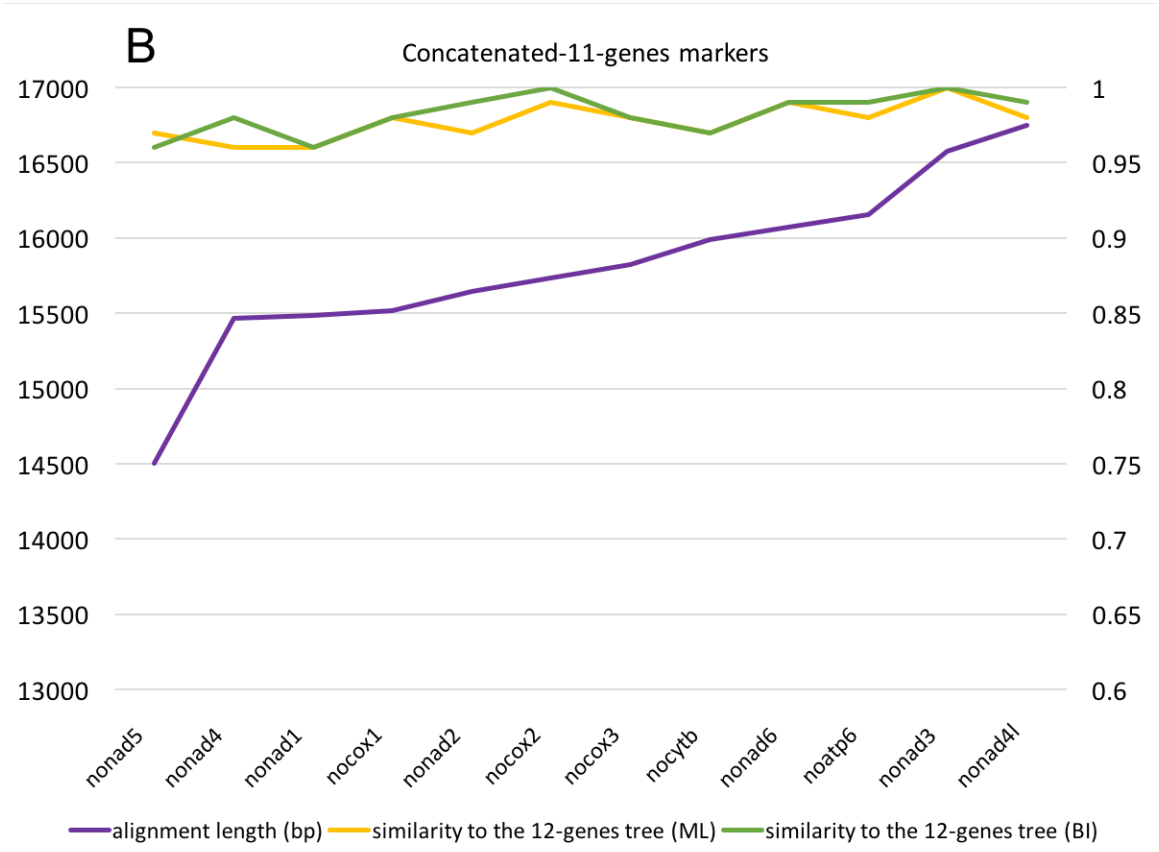

Figure 4.4B. Topological comparison of phylogenetic markers referring to 12-genes tree

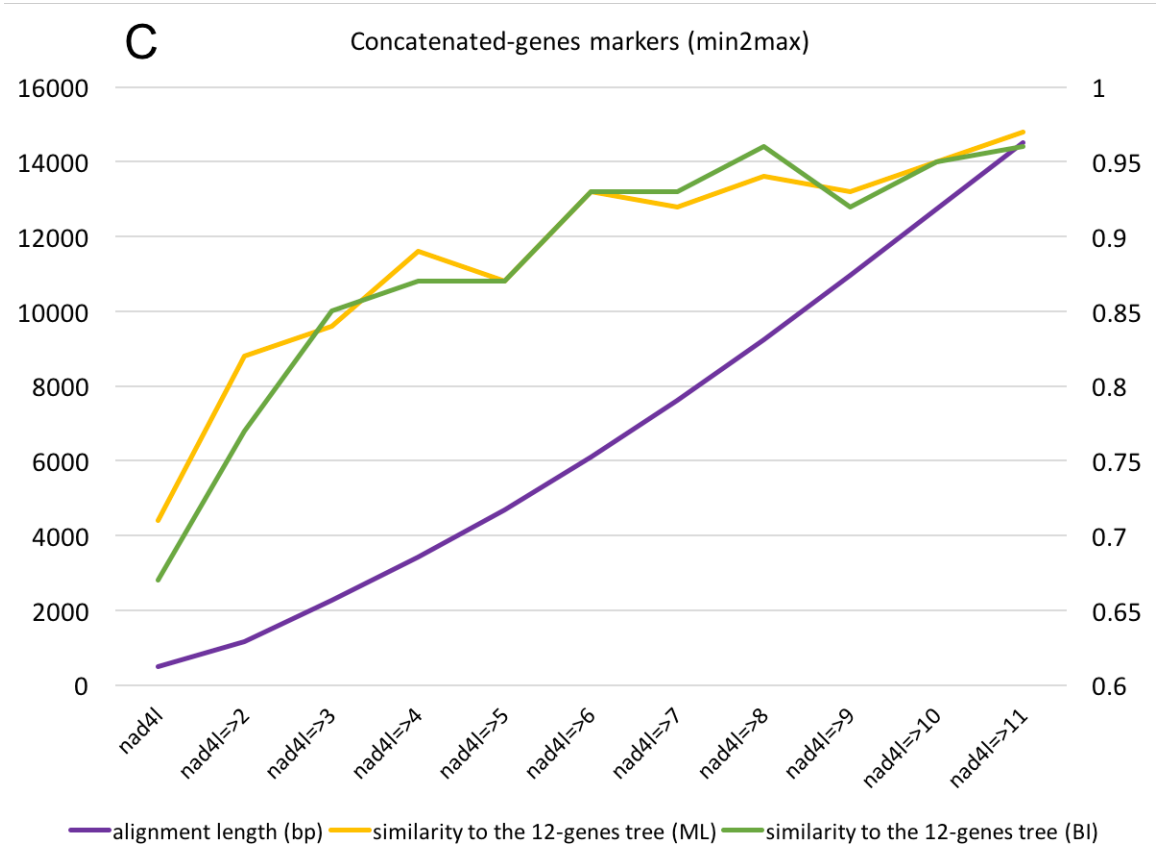

Figure 4.4C. Topological comparison of phylogenetic markers referring to 12-genes tree

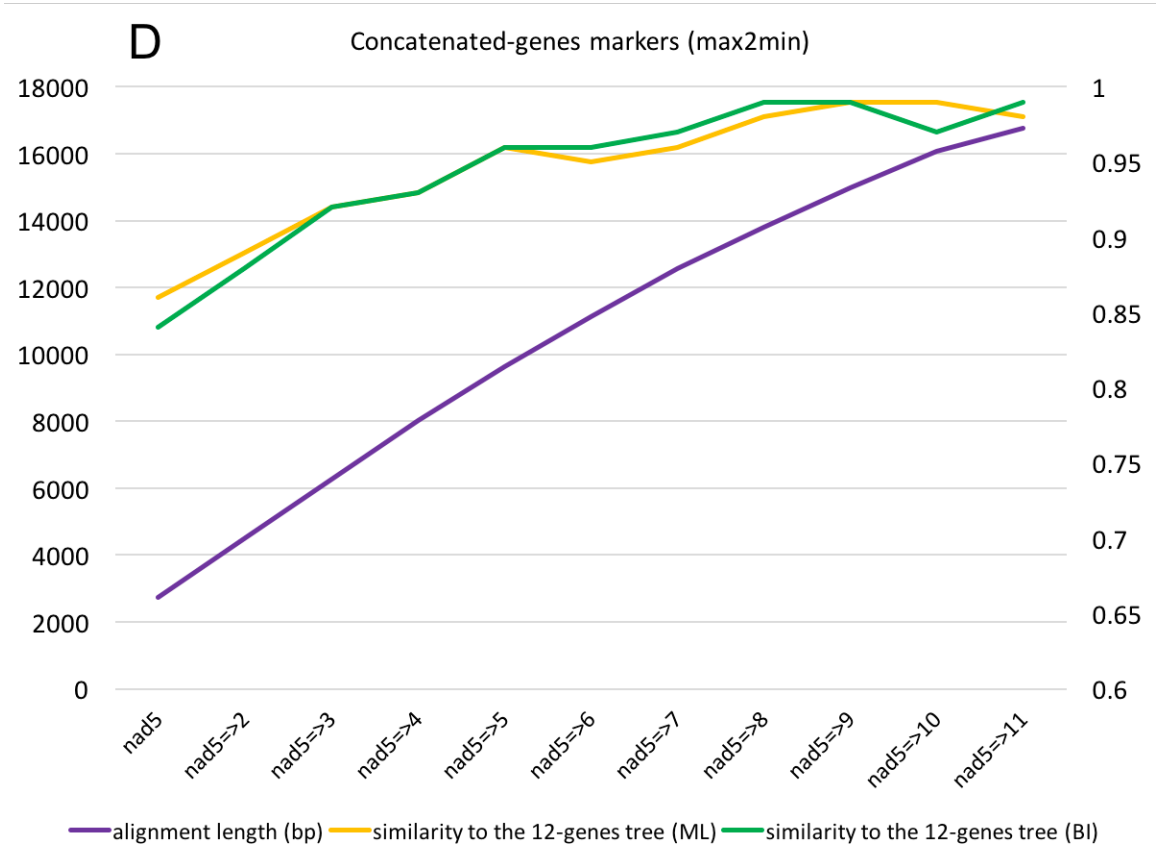

Figure 4.4D. Topological comparison of phylogenetic markers referring to 12-genes tree

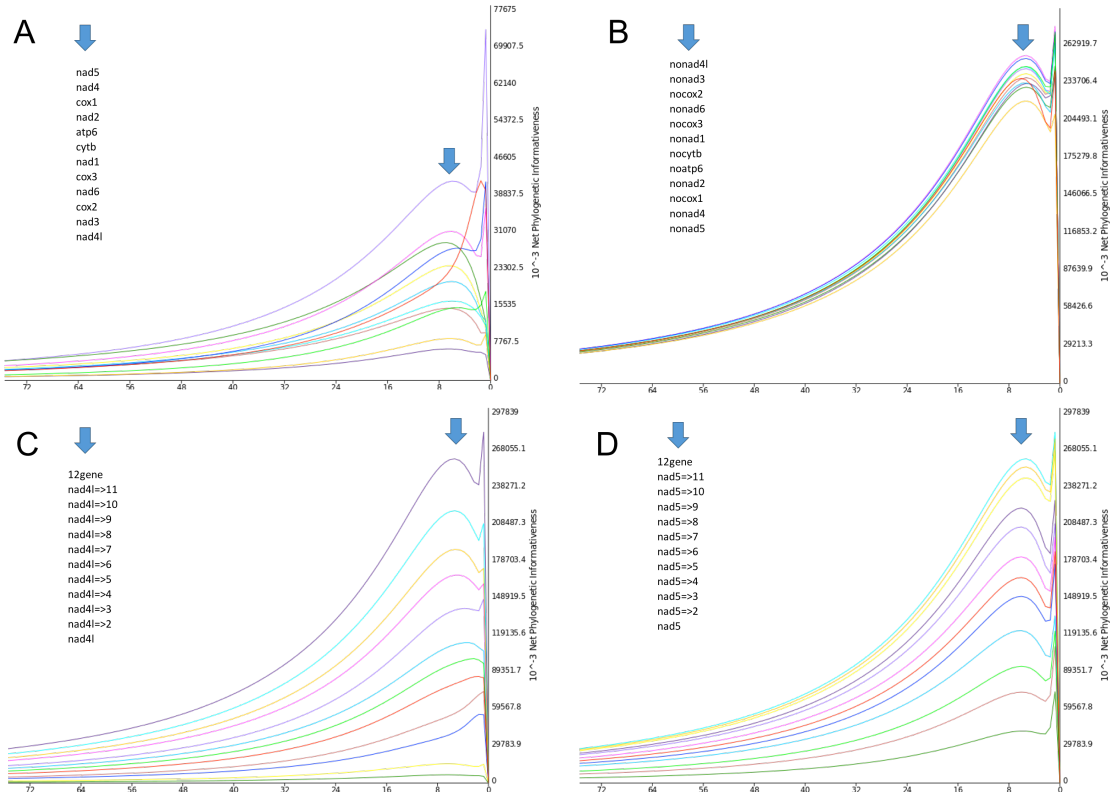

Figure 4.5. Phylogenetic informativeness of all phylogenetic markers studied by PhyDesign. The X axis indicates Time (Million years ago). The Y axis indicates the net phylogenetic informativeness of markers. The arrow indicate the position for ranking markers in each category.

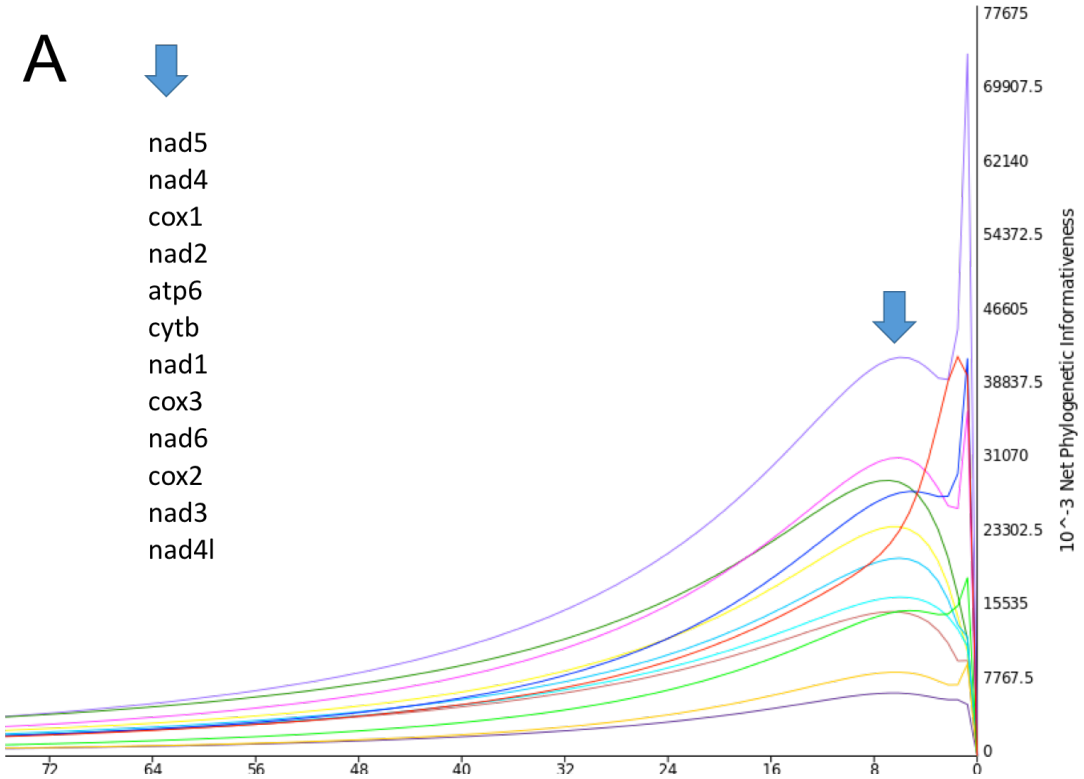

Figure 4.10A Phylogenetic informativeness of all phylogenetic markers studied by PhyDesign. The X axis indicates Time (Million years ago). The Y axis indicates the net phylogenetic informativeness of markers. The arrow indicate the position for ranking markers in each category. A: Single gene markers.

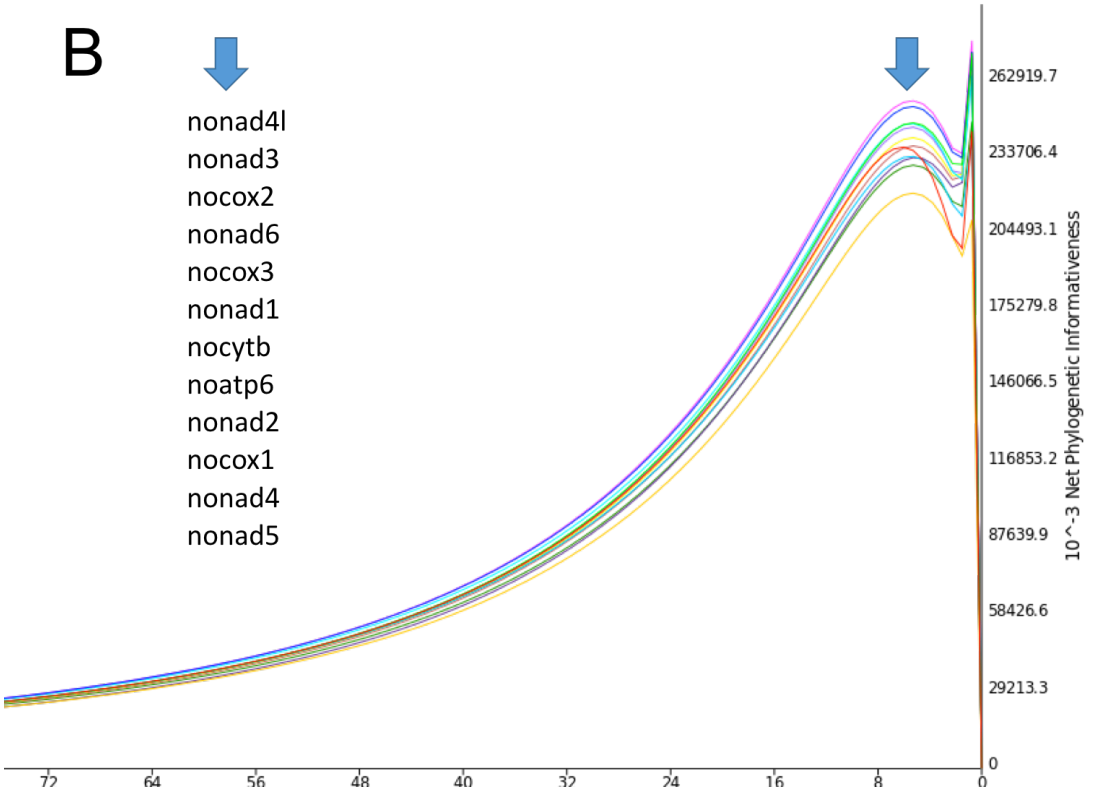

Figure 4.5B. Phylogenetic informativeness of all phylogenetic markers studied by PhyDesign. The X axis indicates Time (Million years ago). The Y axis indicates the net phylogenetic informativeness of markers. The arrow indicate the position for ranking markers in each category. B: 11-genes markers.

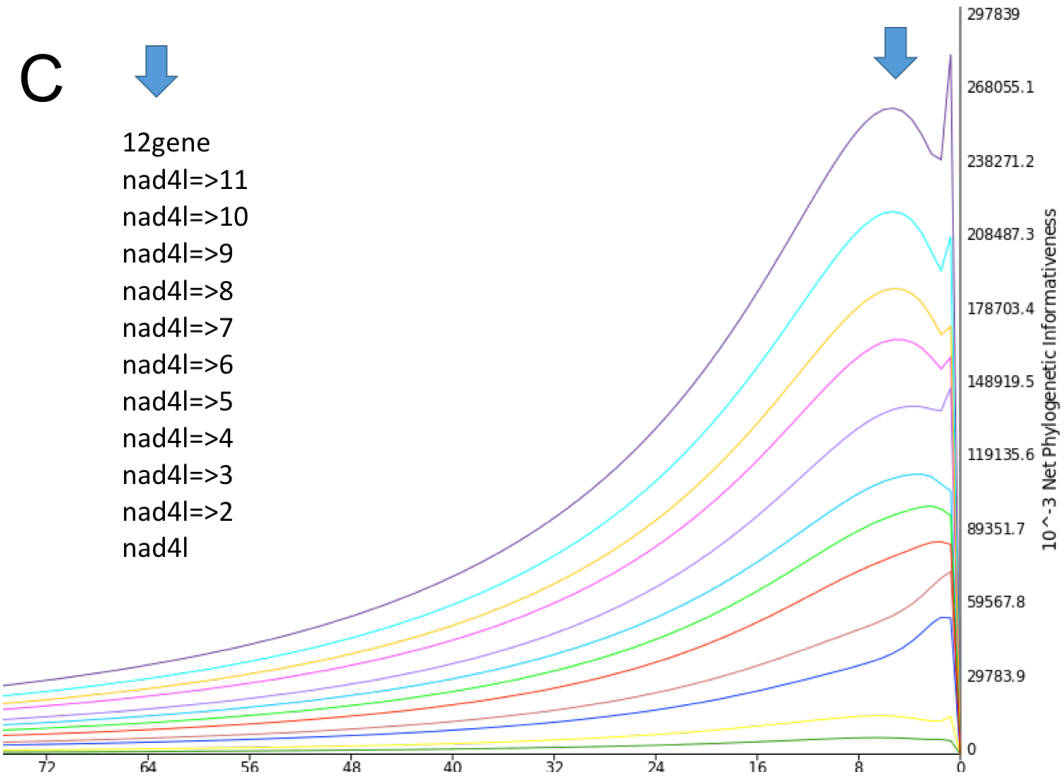

Figure 4.5C. Phylogenetic informativeness of all phylogenetic markers studied by PhyDesign. The X axis indicates Time (Million years ago). The Y axis indicates the net phylogenetic informativeness of markers. The arrow indicate the position for ranking markers in each category. C: Concatenated-genes markers from the smallest single gene.

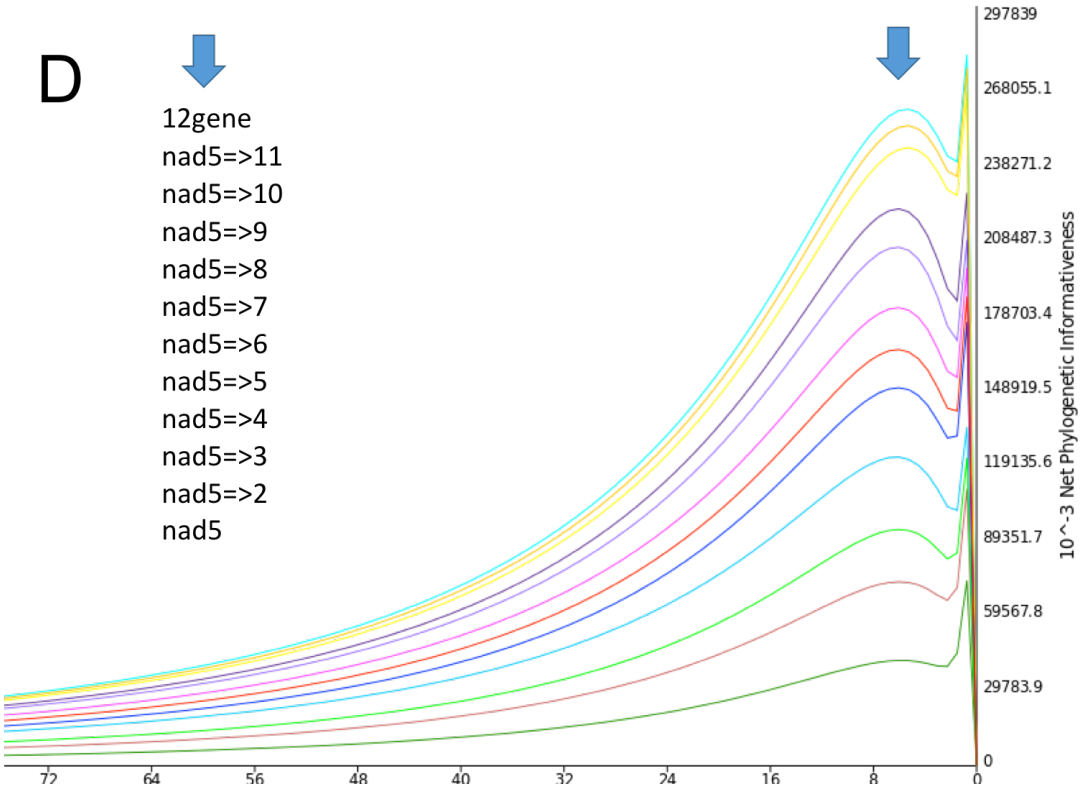

Figure 4.5D. Phylogenetic informativeness of all phylogenetic markers studied by PhyDesign. The X axis indicates Time (Million years ago). The Y axis indicates the net phylogenetic informativeness of markers. The arrow indicate the position for ranking markers in each category. D: Concatenated-genes markers from largest single gene.

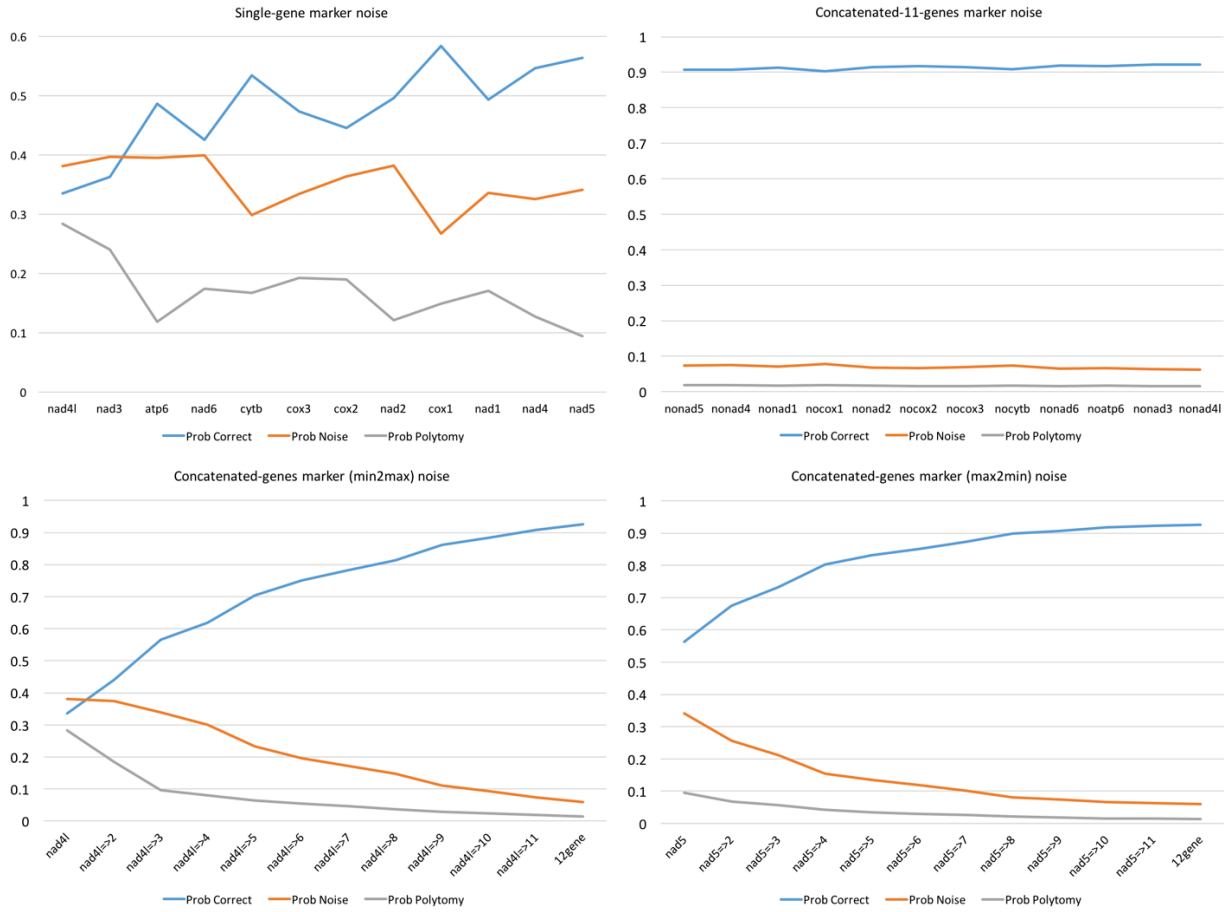

Figure 4.6. Phylogenetic noise analysis of all phylogenetic markers in this study.

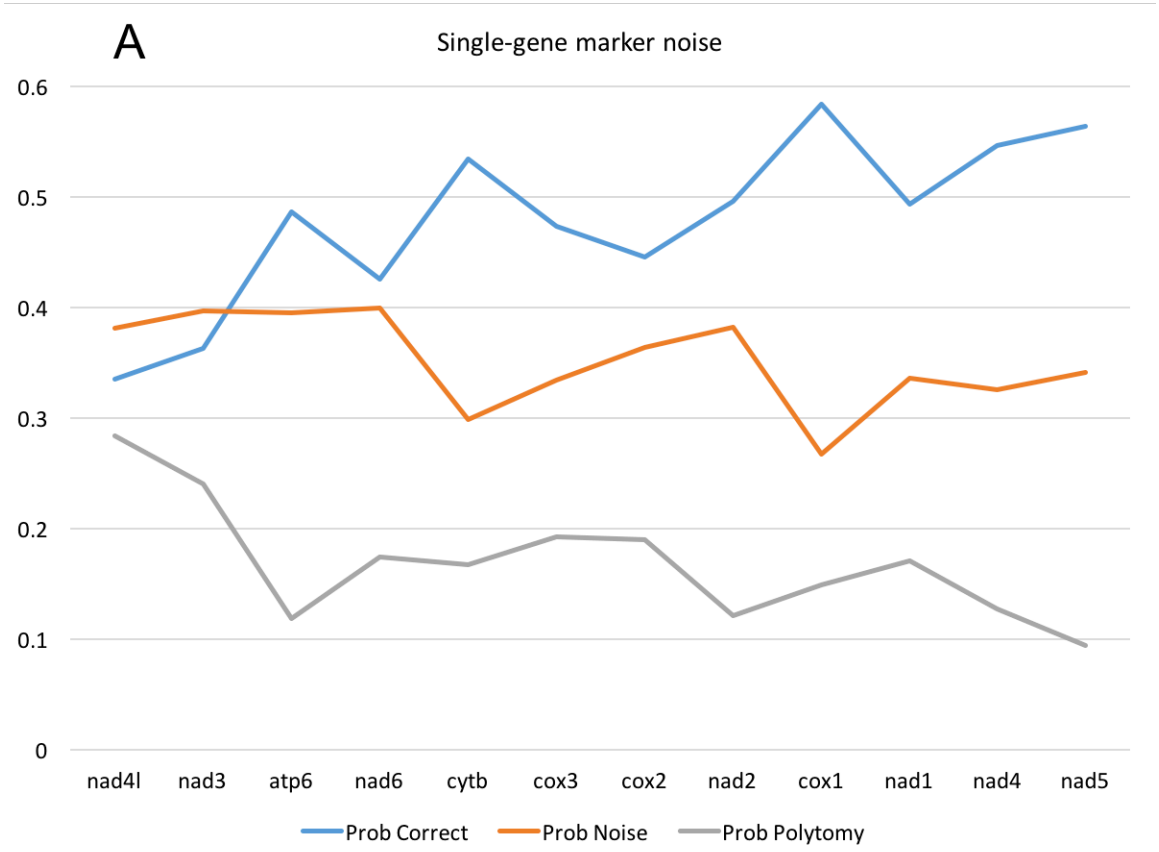

Figure 4.6A. Phylogenetic noise analysis of all phylogenetic markers in this study.

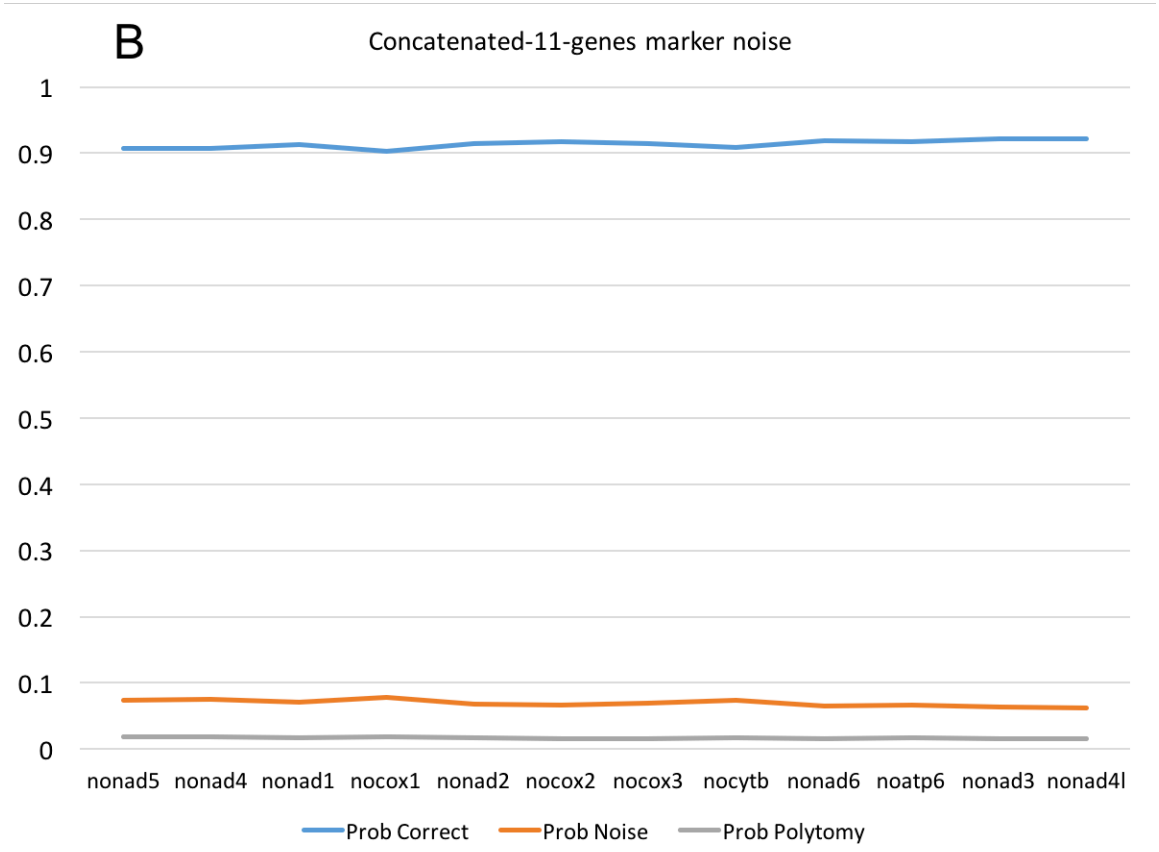

Figure 4.6B. Phylogenetic noise analysis of all phylogenetic markers in this study.

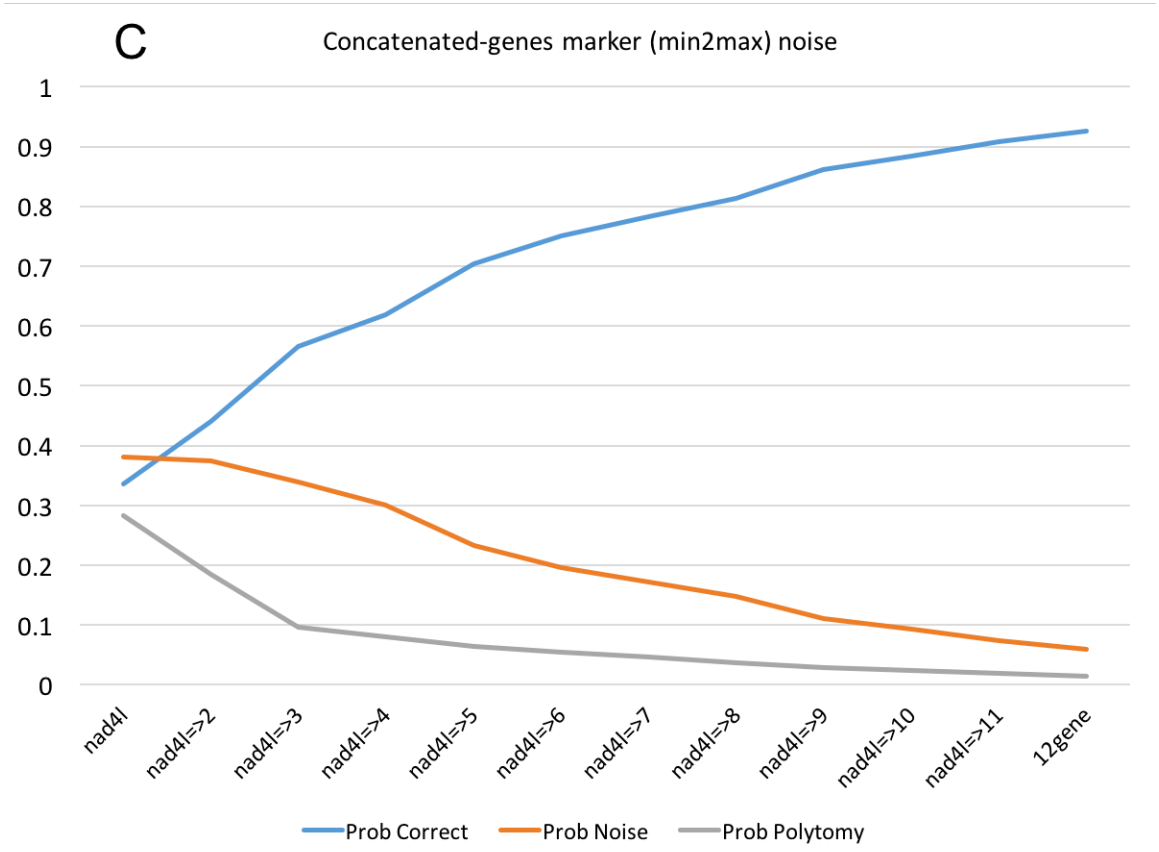

Figure 4.6C. Phylogenetic noise analysis of all phylogenetic markers in this study.

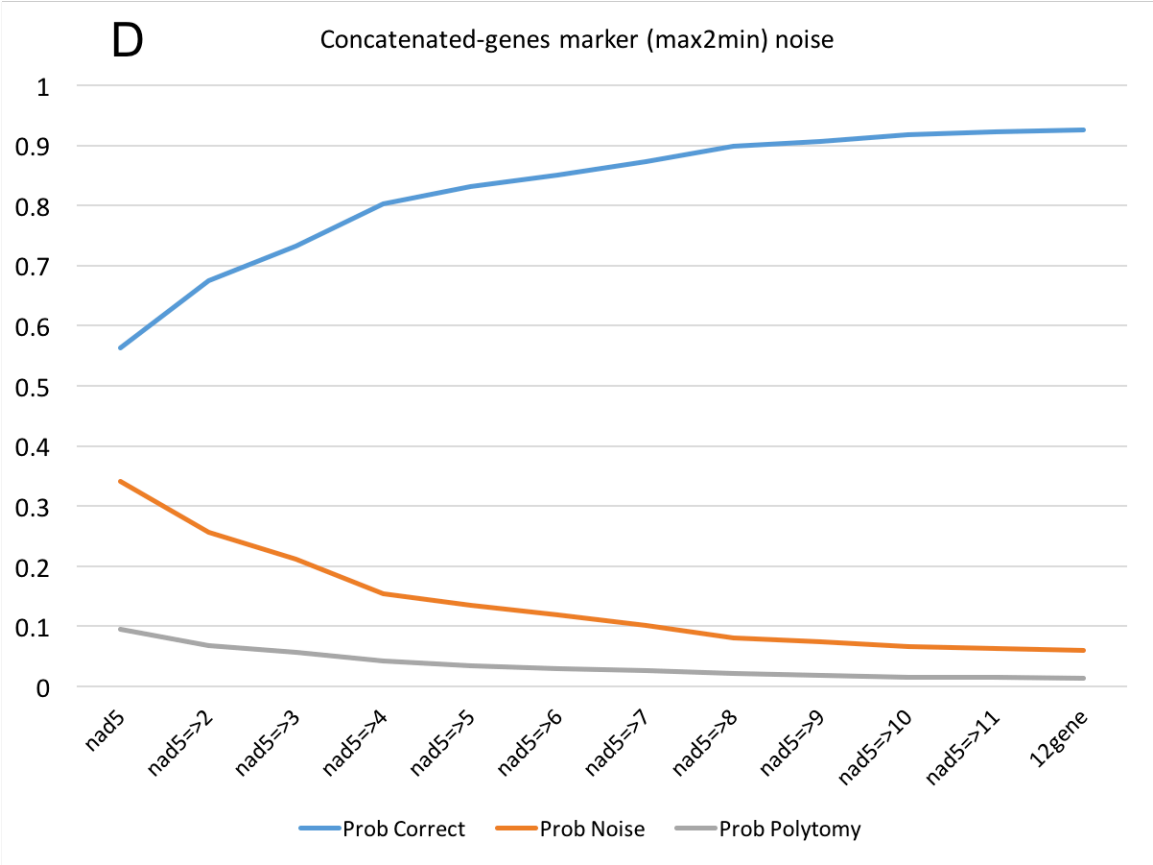

Figure 4.6D. Phylogenetic noise analysis of all phylogenetic markers in this study.

CHAPTER V.

CONCLUSIONS

I. During a survey of fauna of the Great Smoky Mountains National Park, along the Tennessee–North Carolina border in the southeastern United States, a *Hoplolaimus* species was isolated from a mixed forest sample of maple (*Acer* sp.), hemlock (*Tsuga* sp.) and silverbell (*Halesia carolina*). Juveniles, females and males of the lance nematode were examined for morphological characteristics, morphometrics, and phylogenetic relationships, and considered to be an undescribed species. *Hoplolaimus smokyensis* n. sp. is characterized by possession of a lateral field with four incisures, an excretory pore posterior to the hemizonid, esophageal glands with three nuclei, phasmids anterior and posterior to the vulva, and the epiptygma absent. Morphologically, *H. smokyensis* n. sp. is closest to *H. galeatus* and *H. stephanus*, it can be distinguished by minor morphological differences such as the 5-6 annuli in the lip region (compared to 4 in *H. stephanus* or 5 in *H. galeatus*) and morphometric values (average body length longer than *H. stephanus* and shorter than *H. galeatus*). However, *H. smokyensis* n. sp. is considerably distinguished from other species on molecular level. Both nuclear DNA and mitochondrial DNA sequences show significant unique molecular characteristics compared with morphologically similar species. Phylogenetic analyses based on ribosomal and mitochondrial gene sequences also suggest *H. smokyensis* n. sp. to be an independent lineage distinct from all other reported *Hoplolaimus* species.

II. Photos of esophageal gland cells of *H. columbus* in juxtaposition to *H. galeatus* were presented firstly. Both juveniles and adults were observed. Locations of nuclei were identified, confirmed that Siddiqi's statement of four nuclei in the *H. columbus* dorsal gland cell. The anterior and posterior ends of a *H. columbus* male was reported with photographs. Additionally, abnormal female tails were shown, which was mentioned in *H. magnistylus* but first reported in *H. columbus*.

III. The first complete *de novo* assembly of mitochondrial genome (mt-genome) of *H. columbus* is reported using Whole Genome Amplification and Illumina MiSeq technique. The assembly results from multiple methods consent a circularized DNA of 25228bp. This is the largest single mitochondrial chromosome reported in the Tylenchida order. Annotations were performed using invertebrate mitochondrial genetic code 5 and nematode genetic code 14 on MITOS webserver. Both results show 2 ribosomal RNA genes, 12 protein-coding genes, and a large non-coding region over 7000bp. The atp8 gene, genes of trnA and trnM are missing in both results. Results also present discrepancies against each other on PCG loci and tRNA predictions: the code-5-result has a trnI gene and a duplicated gene of trnW but miss trnM gene; on the other side, the code-14-result has a trnM gene and a duplicated trnN gene, but missing trnI. Both annotation results agree on the same PCG order. We also analyzed phylogenetic relationships of 92 taxa nematodes by 12 concatenated protein coding genes, and compared their genome composition, PCG content and order arrangement. Analyses indicate the special

characters of *H. columbus* mitochondrial genome as well as its phylogenetic relationships in the Nematoda phylum.

IV. The phylogenetic informativeness of 93 nematode mitochondrial genomes, including *H. columbus* and *H. galeatus*, were investigated by two quantitative methods: comparison of topological results from diverse phylogenetic markers using ETE3 program, and profiling phylogenetic informativeness and noise of markers using PhyDesign. Two methods agree with each other on phylogenetic signal of nematode mitochondrial protein-coding genes. The results indicate that (1) Phylogenetic informativeness is independent of sequence length. (2) Concatenated-genes marker with suitable genes offer better phylogeny. (3) Nad5 and nad4 contain high phylogeny informativeness as phylogenetic markers of nematodes. (4) Traditional mitochondrial markers, such as the cox1 or ctyb gene, contain medium informativeness for the nematode phylogeny. (5) Nad3, nad4l, cox2 and nad6 genes have low phylogenetic informative, and these genes are not suggested as an individual marker for the nematode phylogeny.

These studies could contribute to future nematological researches as references on taxonomy, phylogeny, and evolutionary biology of nematodes. They could also contribute on lance nematode population study and biogeographical investigation.