DOI: https://doi.org/10.3897/arphapreprints.e61100

COI and ND1 as DNA barcodes to identify *Taenia hydatigena* (Cestoda: Taeniidae) from canine small intestine on Hainan Island in China

Yajun Lu, Siqi Yang, Jingwen Liu, Xixi Lu, Qiuyu Zhao

Title

COI and ND1 as DNA barcodes to identify *Taenia hydatigena* (Cestoda: Taeniidae) from canine small intestine on Hainan Island in China

Author names

Siqi Yang[1], Jingwen Liu[1], Xixi Lu[1], Qiuyu Zhao[1], Yajun Lu[1*]

Author affiliations

1 Key Laboratory of Tropical Translational Medicine of Ministry of Education and School of Tropical Medicine and Laboratory Medicine, Hainan Medical University, Haikou, Hainan 571199, China.

Conflict of interest statement

The co-authors declare that they have no conflict of interest statement.

*Correspondence author

Yajun Lu

E-mail: luyajun@hainmc.edu.cn

Funding

This study was supported by grants from the Natural Science Foundation of Hainan Province (2019RC223).

Ethics approval

Not applicable.

Consent to participate

All participants provided written informed consent.

Consent for publication

All participants obtained written informed consent for publication.

Availability of data and material

The data used to support the findings of this study are included in the article. All data are fully available without restriction.

Code availability

Not applicable.

Authors' contributions

Siqi Yang performed the experiments and wrote the manuscript. Jingwen Liu, Xixi Lu, and Qiuyu Zhao performed the experiments. Yajun Lu designed the study, performed revisions to the draft report, and approved the final manuscript.

Acknowledgements

The authors gratefully thank the Science and Technology Department of Hainan Province for the financial support and Academician Workstation for material support.

Abstract

DNA barcoding based on universal gene markers is a fast, accurate, and innovative approach for the molecular discrimination of species. Some species are particularly difficult to discriminate using a traditional morphological identification method because of severely damaged morphological features. In this study, cytochrome c oxidase subunit I (COI) and NADH dehydrogenase subunit 1 (ND1) were used as barcoding markers to distinguish *Taenia hydatigena* in dogs on the tropical island of Hainan. Therefore, geographic differentiation based on the COI and ND1 sequences amongst the specimens and other geographic isolates in GenBank was determined by calculating the genetic distances according to the Kimura 2-parameter (K2P) model and constructing a phylogenetic tree using the neighbour-joining (NJ) method. Barcoding gap, base composition, and base saturation were tested to assess the effectiveness of the barcoding marker COI and ND1 genes for specimen identification. In addition, we analysed the barcoding gap and saturation and performed molecular evolutionary analysis of the intraspecies and interspecies diversity of Taenia.

Keywords

Cytochrome c oxidase subunit I; DNA barcodes; NADH dehydrogenase subunit 1; *Taenia hydatigena*

Introduction

With the development of molecular biology, it is increasingly common to identify species by combining molecular biology with genetic evolution analysis (Young et al. 2019). Molecular identification based on specific DNA sequence analyses has become a common technological tool used to identify new species and can make more accurate assessments within and between species biodiversity than other methods as well as serve as an effective supplement to traditional morphological identification (Valentini et al. 2009). DNA barcodes have been widely used in biological molecular identification and have great significance for uncovering biological diversity. DNA barcodes can be used to establish a large-scale classification system for biodiversity analysis through several species-specific DNA barcodes (Coissac et al. 2016). DNA barcodes also provide reliable species identification for all life stages and even for tissue debris (Hawlitschek et al. 2016).

The application of the barcode of the COI gene in the study of molecular biodiversity is increasing rapidly and continuously (Balech et al. 2018). There are approximately 4.5 million COI gene barcodes in GenBank and BOLD (Barcode of Life Data Systems), including different individuals across 250,000 species. It has become a common phenomenon to identify different animals by using the barcode of the COI gene (Thaler et al. 2016). In the identification of some species, the COI gene is more suitable for identifying individuals (Laopichienpong et al. 2016). The COI gene may prove to be valuable in the identification of the early developmental stage of species (Jaakko et al. 2016).

T. hydatigena is a zoonotic tapeworm species. Cysticercus is a larva that can be widely parasitized in domestic animals with a high infection rate. Hainan is a high incidence area of parasitic diseases. Taenia is a highly diverse population, and using morphological methods alone may fail to identify some parasites accurately (Ulziijargal et al. 2019). In 2013, Minoru and colleagues constructed molecular phylogeny using mitochondrial genes to explain the genetic relationship among different Taenia (Minoru et al. 2013). Since two gene barcodes have been proven to have better resolution (Pennisi 2007), we use two genes, COI and ND1, as barcodes to reflect the diversity of *T. hydatigena* (Cestoda: Taeniidae) on Hainan Island, and we analyse the molecular evolution of Taenia using molecular genetics.

Materials and methods

Sample collection and genomic DNA extraction

An adult worm was obtained from the canine small intestine. The fresh worm was washed in physiological saline and stored in 70% ethanol until use. Approximately 2-3 millimetres of worm was cut with a scalpel and washed three times with ddH_2O. The small piece of worm tissue was digested in 100 μL Solution AB and 25 μL AD Buffer (Bioteke, Beijing, China) for 10 min at room temperature followed by 5 min at 95 ºC and was subsequently added to 100 μL Solution AC (Bioteke) to extract the genomic DNA.

PCR amplification and sequencing

The COI and ND1 genes were amplified from genomic DNA by polymerase chain

reaction (PCR). The primer sequences for COI were the forward primer 5'-TTTTTTGGGCATCCTGAGGTTTAT-3' and reverse primer 5'-TAAAGAAAGAACATAATGAAAATG-3'. The primer sequences for ND1 were the forward primer 5'-AGATTCGTAAGGGGCCTAATA-3' and reverse primer 5'-ACCACTAACTAATTCACTTTC-3'. Amplification reactions contained 10 μL of 2×Power Taq PCR MasterMix (Bioteke), 4 μL of ddH_2O, 1 μL of each forward and reverse primer (10 μM), and 4 μL of template DNA (0.002 μg/μL). The cycling conditions were as follows: 94 ºC for 3 min, 35 cycles of 94 ºC for 1 min, 55 ºC for 1 min and 72 ºC for 1 min, followed by a final step of 72 ºC for 5 min. PCR products were purified using a FastPure Gel DNA Extraction Mini Kit (Vazyme Biotech, Nanjing, China) and sequenced by the Sanger sequencing method (BGI, Beijing, China).

Molecular evolution analysis

The sequences generated were assembled and refined manually to obtain final high-quality sequences using ApE software v2.0.61 (http://jorgensen.biology.utah.edu/wayned/ape). Multiple sequence alignment was conducted for the sequences of the samples for COI and ND1, the selected sequences of 29 geographic isolates for COI, and 21 geographic isolates for ND1 from the GenBank database using the Basic Local Alignment Search Tool (BLAST) of NCBI (National Center for Biotechnology Information) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The phylogenetic trees were constructed using the neighbour-joining (NJ) algorithm. The intraspecific and interspecific genetic

distances were calculated using the Kimura 2-parameter (K2P) model, and the nucleotide composition, nucleotide frequencies, nucleotide pair frequencies, and transition/transversion ratios were analysed using MEGA-X software.

Haplotype diversity analysis

Haplotype diversity (Hd) was evaluated using DNA Sequence Polymorphism (DNAsp) v6.0 software by calculating the number of segregating sites (S), the average number of nucleotide differences per site (pi), and the total number of mutations.

Effectiveness assessments of the barcoding markers

The barcoding gaps of the markers for COI and ND1 genes were detected using automatic barcode gap discovery (ABGD). The values of substitution saturation were calculated using Data Analysis in Molecular Biology and Evolution (DAMBE) v7.2.14 software.

Results

Morphological characters

The adult worm was opalescent-yellowish and measured 122.4 cm in length. The strobila had 250-300 proglottids. The immature proglottids were wider than longer, and the mature proglottids were longer than wide at 10-14 mm long and 4-5 mm wide. A mature proglottid had an ovary and 600-700 testes. The uterus, filled with eggs, had 5-10 uterine branches with smaller branches in a gravid proglottid.

Each genital pore was unilateral and irregularly alternating. The eggs were oval in

shape and approximately 36-39 mm in size, containing an oncosphere (Figure 1).

Phylogenetic characteristics

A 422-bp fragment of the COI gene was successfully amplified, and a 385-bp fragment was retained after manual editing. Additionally, a 502-bp fragment of the COI gene was successfully amplified, and a 471-bp fragment was retained after manual editing. The phylogenetic tree of the COI sequence (Figure 2) showed that *T. hydatigena* Hainan isolates had the closest relationship with *T. hydatigena* from Iran (Table 1), and the phylogenetic tree of the ND1 sequence (Figure 3) showed that *T. hydatigena* Hainan isolates had the closest relationship with *T. hydatigena* from Kenya (Table 2). The nucleotide sequence homology among the *Taenia hydatigera* was very similar, showing high homogeneity. For the ND1 gene, the genetic relationship between *Taenia hydatigera* and *Taenia saginata* was the furthest.

Genetic diversity analyses

A total of 15 haplotypes were obtained from the sequence alignment of 29 isolates based on the COI gene: the haplotype diversity was 0.8276, the nucleotide diversity pi was 0.07769, and the number of segregating sites was 130. The average content of the A, T, C, and G bases from the COI gene sequences was 22.3%, 44.7%, 9.6%, and 23.5%, respectively, and the GC content was 33.1%. The COI gene contained 81 variable sites. The transitional pairs (si) were 10, and the transversional pairs (sv) were 13 with a si/sv ratio of 0.77.

A total of 15 haplotypes were obtained from the sequence alignment of 21 isolates

based on the ND1 gene: the haplotype diversity was 0.9381, the nucleotide diversity pi was 0.12544, and the number of segregating sites was 260. The average content of A, T, C, and G bases from ND1 gene sequences was 19.7%, 51.6%, 6.5%, and 22.3%, respectively, and the GC content was 28.8%. The ND1 gene contained 94 variable sites. The transitional pairs (si) were 10, and the transversional pairs (sv) were 11 with a si/sv ratio of 0.91. The transversional pairs of the COI gene turned into a straight upward trend with increasing genetic distance, and when the base conversion proportion was 3%, the platform period appeared. For ND1, both transitional pairs and transversional pairs had a platform period. The mutation of the ND1 gene reached a saturation state (Figure 4).

Effectiveness of the barcoding markers

The COI gene had obvious DNA barcoding gaps. The DNA barcoding gaps of the COI gene were 0.03-0.12 and 0.20-0.30, and the DNA barcoding gap of the ND1 gene was 0.45-1.25 (Figure 5). The COI and ND1 genes both showed DNA barcoding gaps within and between populations, and they had a certain genetic interval; therefore, they could be used to distinguish species.

We used MEGA-X to calculate the genetic distance within and between different tapeworms in COI and ND1. For COI, the maximum genetic distance was between *Taenia saginata* and *Taenia taeniaformis*. The genetic distance between *Taenia saginata* and other worms was far. For ND1, the results showed that the genetic distance between *Taenia hydatigera* and *Taenia saginata* was the largest, and they had

the furthest genetic relationship. The histogram of distance shows that the COI sequence set is divided into 8 distinct groups, and the ND1 sequence set is divided into 7 distinct groups.

Discussion

T. hydatigera is an endoparasite that causes human and animal infections. Adult worms inhabit the small intestine of canines, wolves, and other canids. The larvae disseminate to the livers, lungs, and various other organs of livestock and wildlife. Humans usually develop infections from ingesting food and water contaminated by parasite eggs in canine faeces. In 2003, Hebert et al. first proposed the term DNA barcoding (Hebert et al. 2003) as a technique to identify organisms. DNA barcoding is an efficient tool for distinct species (DeSalle et al. 2005; Frézal and Leblois 2008). Accumulating evidence shows that the COI gene can serve as the core of a global bioidentification system for animals. The COI gene can distinguish species at the DNA level. A DNA barcoding technique based on the molecular evolution of the COI gene is a reliable classification method. The COI gene is a general DNA barcoding marker; nevertheless, the molecular identification rate of sibling species or related species of ecogeographical groups is relatively low. In addition, DNA barcoding markers of multiple genes have better discriminability than a single gene as a DNA barcoding marker. An outstanding DNA barcoding marker requires that there is an obvious barcoding gap between intra- and interspecific distances. The average interspecific distances were more than 10-fold larger than the average intraspecific distances (Paul et al. 2004). Therefore, we used the COI and ND1 genes as DNA

barcoding markers to identify the species of the worm and analysed the phylogenetic relationships of the specimen and different geographic isolates in this study. The *T. hydatigena* Hainan isolate had the closest phylogenetic relationship with the *T. hydatigena* Chabahar isolate from Iran based on the COI gene. The *T. hydatigena* Hainan isolate had the closest phylogenetic relationship with the *T. saginata* Kenya isolate from Iran based on the ND1 gene. The phylogenetic tree and genetic distance showed that the genetic distance between *Taenia saginata* and other worms was far. Two different genes had the same result. Through the existence of a barcode gap, we could automatically classify species sequences. The existence of a barcode gap proved that species can be distinguished effectively (Puillandre et al. 2011). The COI gene and the ND1 gene both showed an obvious barcode gap. There were two barcode gaps in the COI gene. These results suggest that the identification effect of the COI gene is better in species and interspecies of Taenia.

After the analysis of bases, we found that the haplotype diversity, base polymorphism, and polymorphic isolate sites of the ND1 gene were all high. In addition, the mutation of the ND1 gene reached a saturation state. Mitochondrial genes often have mutation saturation and a rapid mutation rate, so in the process of species formation, a base site may have multiple mutations, which will lead to a reduction in genetic distance when comparing distant species (Blouin et al. 1998). Using the COI gene may be more accurate and suitable.

In conclusion, it is extremely important to use DNA barcodes for species identification, especially for species that are morphologically similar and difficult to

distinguish by morphological methods. At present, due to the wide application of DNA barcodes, many researchers have conducted research on DNA barcodes and have proposed methods with high precision and calculation efficiency to improve the classification and recognition of barcodes (Goldstein and DeSalle 2011). DNA barcode identification of species has a good perspective. The principle of DNA barcodes is to select highly conserved DNA coding regions or noncoding region fragments with subtle variation at the evolutionary level to identify species, but it is difficult to find such general standard sequence fragments. In the process of species identification, we should also consider the mutation saturation of bases and properly use the genetic information analysis of base transversion to improve accuracy. Our study is relevant for its identification of *T. hydatigena* in Hainan. Moreover, with the development of DNA barcode research, DNA barcode technology will significantly advance the biodiversity research field.

References

Balech B, Sandionigi A, Manzari C, Trucchi E, Saverio V (2018) Tackling critical parameters in metazoan meta-barcoding experiments: A preliminary study based on coxI DNA barcode. PeerJ 6: e4845. https://doi.org/10.7717/peerj.4845

Blouin MS, Yowell CA, Courtney CH, Dame JB (1998) Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. Molecular Biology & Evolution 12: 1719-1727. https://doi.org/10.1093/oxfordjournals.molbev.a025898

247 Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to
248 genomes: extending the concept of DNA barcoding. Molecular Ecology 25:
249 1423-1428. https://doi.org/10.1111/mec.13549
250 DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species
251 delimitation and DNA barcoding. Philosophical Transactions of the Royal
252 Society B Biological Sciences 360: 1905-1916.
253 https://doi.org/10.1098/rstb.2005.1722
254 Frézal L, Leblois R (2008) Four years of DNA barcoding: current advances and
255 prospects. Infection Genetics & Evolution 8: 727-736.
256 https://doi.org/10.1016/j.meegid.2008.05.005
257 Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic
258 practice: determination, discovery, and description. Bioessays 33: 135-147.
259 https://doi.org/10.1002/bies.201000036
260 Hawlitschek O, Morinière J, Dunz A, Franzen M, Rödder D, Glaw F, Haszprunar G
261 (2016) Comprehensive DNA barcoding of the herpetofauna of Germany.
262 Molecular Ecology Resources 16: 242-253.
263 https://doi.org/10.1111/1755-0998.12416
264 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications
265 through DNA barcodes. Proceedings Biological sciences 270: 313-321.
266 https://doi.org/10.1098/rspb.2002.2218
267 Jaakko LOP, Kahanpää J, Mutanen M (2016) DNA Barcodes for the Northern
268 European Tachinid Flies (Diptera: Tachinidae). Plos One 11: e0164933.

https://doi.org/10.1371/journal.pone.0164933

Laopichienpong N, Muangmai N, Supikamolseni A, Twilprawat P, Chanhome L, Suntrarachun S, Peyachoknagul S, Srikulnath K (2016) Assessment of snake DNA barcodes based on mitochondrial COI and Cytb genes revealed multiple putative cryptic species in Thailand. Gene 594: 238-247. https://doi.org/10.1016/j.gene.2016.09.017

Minoru, Nakao, and, Antti, Lavikainen, and, Takashi, Iwaki, and, Voitto (2013) Molecular phylogeny of the genus Taenia (Cestoda: Taeniidae): Proposals for the resurrection of Hydatigera Lamarck, 1816 and the creation of a new genus Versteria. International Journal for Parasitology 43: 427-437. https://doi.org/10.1016/j.ijpara.2012.11.014

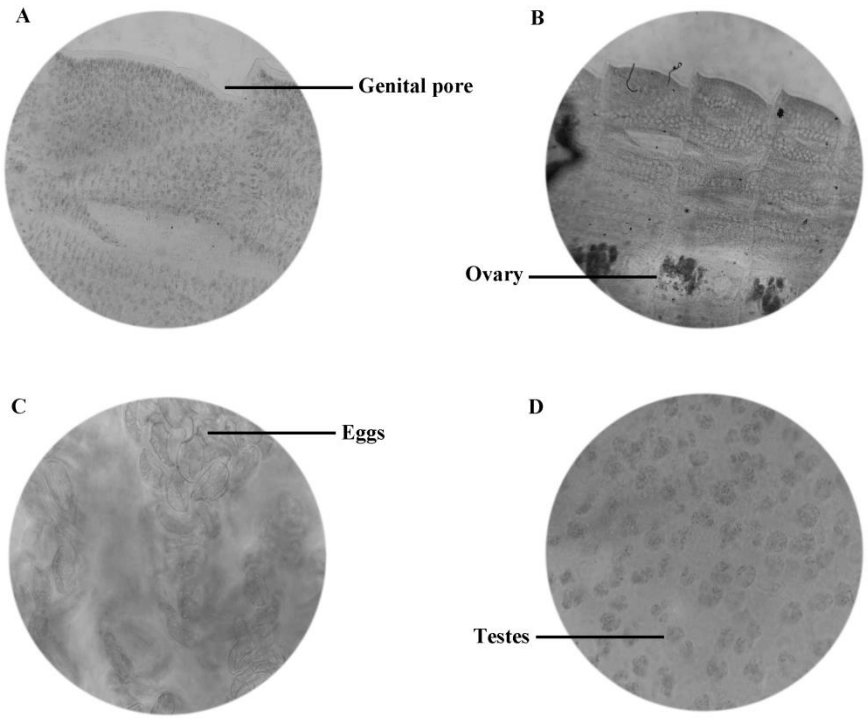
Paul DNH, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten Species in One: DNA Barcoding Reveals Cryptic Species in the Neotropical Skipper Butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences of the United States of America 101: 14812-14817. https://doi.org/10.1073/pnas.0406166101

Pennisi E (2007) Taxonomy. Wanted: a barcode for plants. Science 318: 190-191. https://doi.org/10.1126/science.318.5848.190

Puillandre N, Lambert A, Brouillet S, Achaz G (2011) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular Ecology 21: 1864-1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

Thaler, David S, Stoeckle, Mark Y (2016) Bridging two scholarly islands enriches

291 both: COI DNA barcodes for species identification versus human
292 mitochondrial variation for the study of migrations and pathologies. Ecology
293 and Evolution 6: 6824-6835. https://doi.org/10.1002/ece3.2394
294 Ulziijargal G, Yeruult C, Khulan J, Gantsetseg C, Wandra T, Yamasaki H,
295 Myadagsuren N (2019) Molecular identification of Taenia hydatigena and
296 Mesocestoides species based on copro-DNA analysis of wild carnivores in
297 Mongolia. International Journal for Parasitology: Parasites and Wildlife 11:
298 72-82. https://doi.org/10.1016/j.ijppaw.2019.12.004
299 Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. Trends in
300 Ecology & Evolution 24: 110-117. https://doi.org/10.1016/j.tree.2008.09.011
301 Young MR, Proctor HC, deWaard JR, Hebert PDN (2019) DNA barcodes expose
302 unexpected diversity in Canadian mites. Molecular Ecology 28: 5347-5359.
303 https://doi.org/10.1111/mec.15292



304

305 Figure 1. *T. hydatigena* tissues were made transparent in xylene.

306 The microscope magnification is ×100 in 1-A, C, and D.

307 The microscope magnification is ×40 in 1-B.

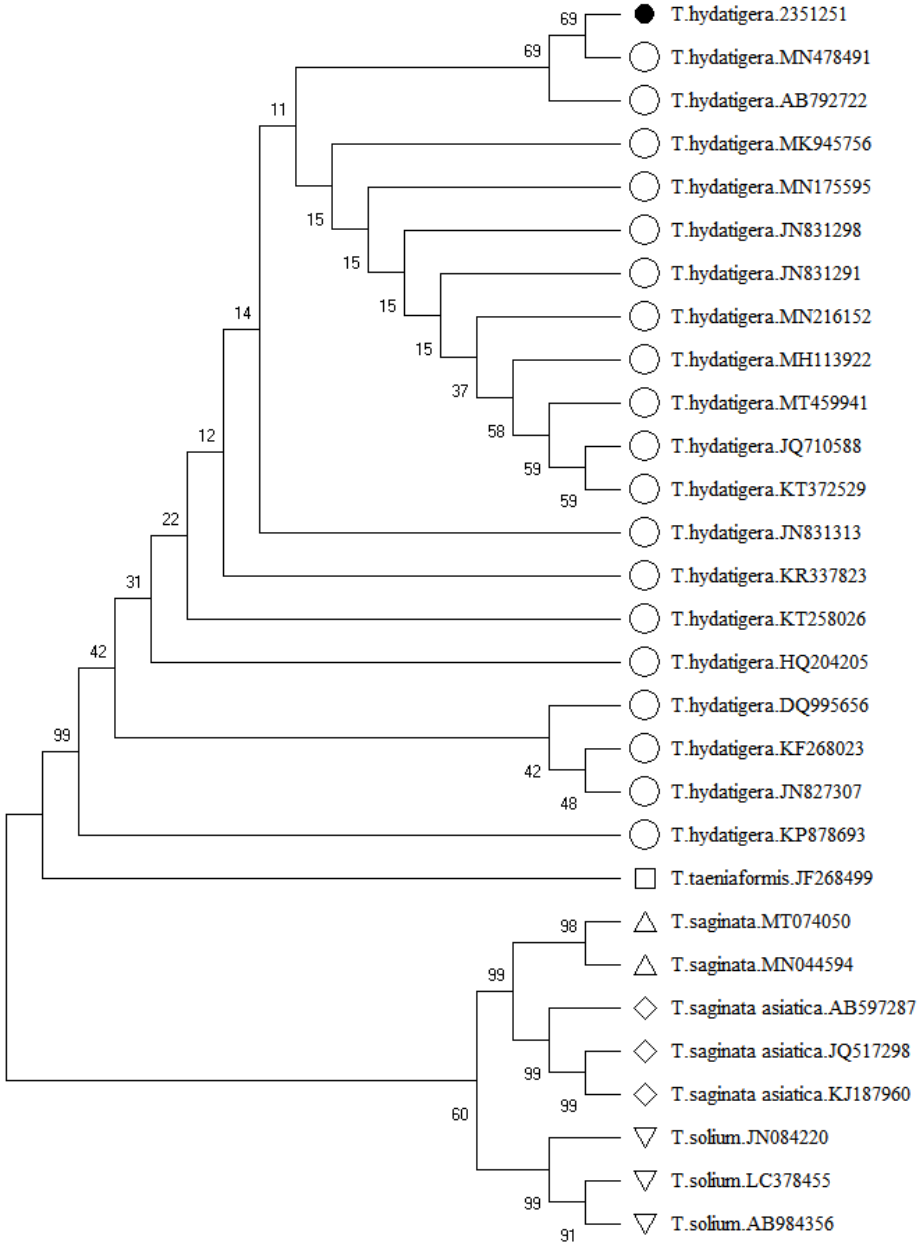
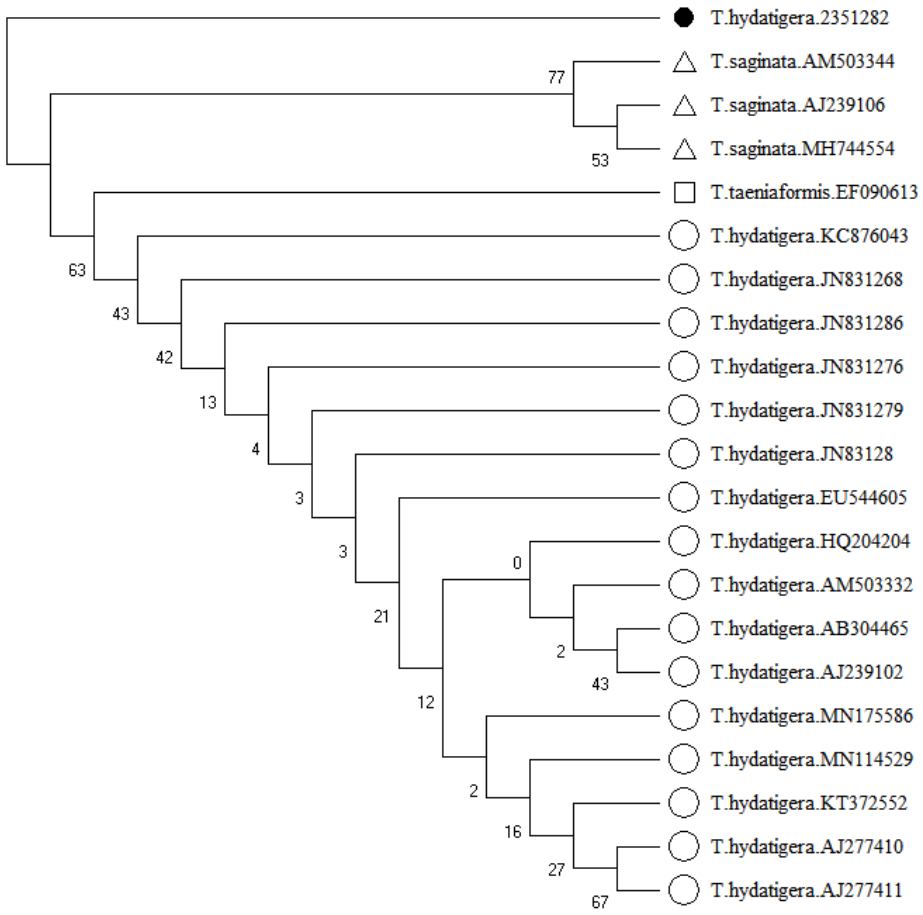


Figure 2. The COI NJ phylogenetic tree is shown for 29 different geographic isolates. Different symbols indicate different species. There are five different kinds of *Taenia spp*. The black circle denotes the *T. hydatigena* (Cestoda: Taeniidae) in Hainan identified by the COI gene.



313

314 Figure 3. The ND1 NJ phylogenetic tree is shown for 21 different geographic isolates.

315 Different symbols indicate different species. There are three different kinds of *Taenia*

316 *spp*. The black circle denotes the *T. hydatigena* in Hainan identified by the ND1 gene.

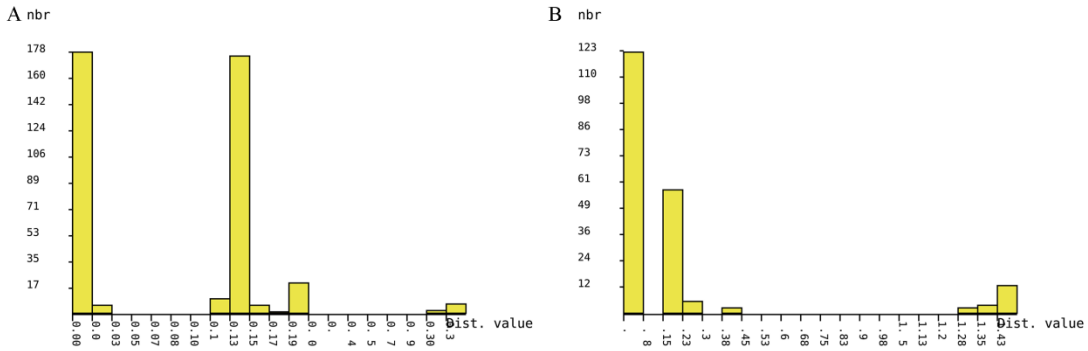


Figure 4. Results of barcoding gap analysis of 29 COI genes (A) and 21 ND1 genes (B) generated by ABDG. The abscissa represents the distance value, and the ordinate represents the numbers. Both genes showed an obvious barcoding gap.

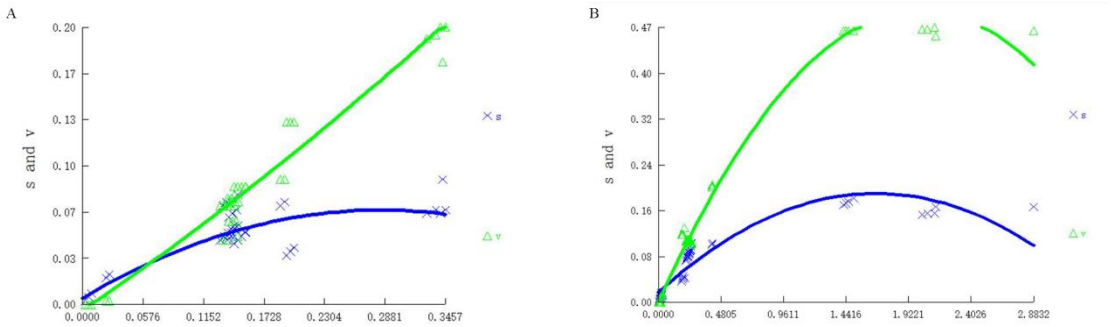


Figure 5. Results of base saturation analysis of 29 COI genes (A) and 21 ND1 genes (B). The abscissa indicates genetic distance, and the ordinate indicates base substitution frequency.

Table 1. The COI gene of the 29 different geographic isolates

GenBank ID	Species	Source	Query Cover (%)	Per.Ident(%)
MN478491	*Taenia hydatigera*	Chabahar	98	99.52
AB792722	*Taenia hydatigera*	Mongolia	98	99.52
HQ204205	*Taenia hydatigera*	China	98	99.28
DQ995656	*Taenia hydatigera*	India	98	99.28
MN216152	*Taenia hydatigera*	*Zambia*	96	99.26
MK945756	*Taenia hydatigera*	Ghana	98	99.04
MN175595	*Taenia hydatigera*	Nigeria	98	99.04
JN831298	*Taenia hydatigera*	China	98	99.04
JN831291	*Taenia hydatigera*	China	98	99.04
MH113922	*Taenia hydatigera*	Iraq	93	99.75
KR337823	*Taenia hydatigera*	Kashmir	98	98.80
JN831313	*Taenia hydatigera*	China	98	98.80
MT459941	*Taenia hydatigera*	Pakistan	91	100
KT258027	*Taenia hydatigera*	China	95	99.26
JQ710588	*Taenia hydatigera*	Iran	91	100
KT372529	*Taenia hydatigera*	Italy	91	100
KP878693	*Taenia hydatigera*	Peru	94	99.25
KF268023	*Taenia hydatigera*	Turkey	97	99.03
JN827307	*Taenia hydatigera*	Turkey	95	99.26
JN084220	*Taenia solium*	Africa	98	87.08
LC378455	*Taenia solium*	Indonesia	96	86.80
AB984356	*Taenia solium*	China	96	86.80
MN337881	*Taenia saginata*	Slovakia	97	87.38
MT074050	*Taenia saginata*	Cambodia	96	87.50
MN044594	*Taenia* saginata	China	98	87.05
AB597287	*Taenia saginata asiatica*	Japan	96	86.76
JQ517298	*Taenia saginata asiatica*	Thailand	96	86.76

KJ187960	*Taenia saginata asiatica*	India	92	86.48
JF268499	*Taenia taeniaformis*	Slovakia	88	87.67

Table 2 The ND1 gene of the 21 different geographic isolates

GenBank ID	Species	Source	Query Cover (%)	Per.Ident (%)
HQ204204	*Taenia hydatigera*	China	90	98.90
JN831281	*Taenia hydatigera*	China	90	98.89
JN831279	*Taenia hydatigera*	China	90	98.89
MN175586	*Taenia hydatigera*	Nigeria	90	98.67
KC876043	*Taenia hydatigera*	Turkey	90	98.46
JN831286	*Taenia hydatigera*	China	90	98.67
JN831276	*Taenia hydatigera*	China	90	98.67
JN831268	*Taenia hydatigera*	China	90	98.45
AM503332	*Taenia hydatigera*	Kenya	86	99.08
MN114529	*Taenia hydatigera*	Sudan	85	99.07
AB304465	*Taenia hydatigera*	Japan	86	99.85
EU544605	*Taenia hydatigera*	Finland	86	99.85
AJ239102	*Taenia hydatigera*	the United Kingdom	84	97.40
KT372552	*Taenia hydatigera*	Italy	77	98.97
AJ277410	*Taenia hydatigera*	Poland	100	97.43
AJ277411	*Taenia hydatigera*	Ukraine	100	88.43
AM503344	*Taenia saginata*	Kenya	79	81.70
AJ239106	*Taenia saginata*	Australia	76	80.88
MH744554	*Taenia saginata*	Slovakia	70	81.41
EF090613	*Taenia taeniaformis*	India	82	98.06