Review

Chinese Science Bulletin

Genomics

January 2012 Vol.57 No.1: 54–67
doi: 10.1007/s11434-011-4748-9

# Advances in the study of helminth mitochondrial genomes and their associated applications

JIA WanZhong*, YAN HongBin, NI XingWei, LOU ZhongZi, LI HongMin, CAO Ping & CAI XuePeng

*State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Key Laboratory of Veterinary Public Health of Ministry of Agriculture, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046, China*

Received February 16, 2011; accepted June 15, 2011

Helminths, including flatworms and roundworms, are abundant organisms that have a variety of life histories. Of these, the genera *Schistosoma*, *Echinococcus*, *Trichinella* are notable parasites of veterinary and medical importance, and cause substantial socio-economic losses throughout China and the rest of the world. Genetic markers in the mitochondrial (mt) genome have proven useful for systematic, ecological, evolutionary and population studies, and the growth of mt genomic research has increased in the last two decades. Technological improvements, such as the long-polymerase chain reaction method and high-throughput sequencing have allowed minute amounts of DNA from single worms, biopsy samples or microscopic organisms to be used for whole mt genome characterization. To facilitate the retrieval, annotation and analyses of mitochondrial features, multiple databases and specific software have also been designed and established. This review focuses on current progress, applications and perspectives regarding helminth mt genomics. To date, the complete mt genomes for 93 species of helminths have been sequenced and analyzed. Analyses of the mt genes, including gene content, arrangement, composition and variation have revealed unique features among the helminths when compared with other metazoans. This provides important data concerning their functional and comparative mitochondrial genomics, molecular taxonomy and characterization, population genetics and systematics, and evolutionary history. Moreover, mt genome data for parasitic helminths are important for diagnosis, epidemiology and ecology of infections. Mitochondrial genome data offer a rich source of markers for the systematics and population genetics of socioeconomically important parasitic helminths of humans and other animals.

**helminths, mitochondrial genome, evolution, molecular taxonomy, DNA barcoding**

**Citation:** Jia W Z, Yan H B, Ni X W, et al. Advances in the study of helminth mitochondrial genomes and their associated applications. Chin Sci Bull, 2012, 57: 54–67, doi: 10.1007/s11434-011-4748-9

Helminths include parasitic worms that belong to a diverse group of animal phyla, including Platyhelminthes, Nematoda, Nematomorpha and Acanthocephala. By far the most numerous and diverse of the helminths are the parasitic flatworms and roundworms, comprising species with remarkably different life histories. Many helminths species are pathogens of animals, including humans, and even plants, and cause significant diseases and major socio-economic losses [1–4]. Central to the control of these parasite infections is knowledge of their population genetics, which also has critical implications for understanding transmission patterns, ecology and evolution [5–8]. Although common and ubiquitous eukaryote molecular markers such as the 18S ribosomal ribonucleic acid (rRNA) gene (18S ribosomal deoxyribonucleic acid (rDNA)) and internal transcribed spacers (ITS), including the ITS1 and ITS2, can be useful and reliable for the identification of helminths species, mitochondrial (mt) genes are also popular and perhaps more effective and reliable markers for deoxyribonucleic acid (DNA) barcoding, offering greater sequence variation and the ability to differentiate closely related species. Mitochondria are sub-cellular organelles involved in oxida-

*Corresponding author (email: jwanzhong@yahoo.com.cn)

tive phosphorylation, which provide energy to the cells within an organism. They play crucial roles in cellular metabolism, survival and apoptosis. Within metazoans, helminth species possess a compact, circular mt genome ranging from 14 to 20 kb in size. Due to the high mutation rates and proposed maternal inheritance of mtDNAs for helminths, they have been considered to be particularly applicable to population genetics, systematic and evolutionary investigations, and for taxonomic identification, among others [9,10].

## 1 Helminth species or strains with entire mt genomes sequenced to date

To date, 27, 16, 1, 46 and 1 entire mt genomes from cestodes, trematodes, planaria, nematodes and acanthocephalans, respectively, have been sequenced according to the NCBI GenBank database and published reports since the mt genomes of *Ascaris suum* and *Caenorhabditis elegans* were characterized in 1990 (Table 1) (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/taxtree.cgi?db=Mito&taxid=33208&result=frame&complete=All&init_rankid=5) [11].

## 2 Methods for investigating complete helminth mt genome sequences

Main procedures for sequencing and analyses of helminths mt genomes include preparation of worm samples, isolation of total genomic DNA from these worms, amplification of overlapping fragments of mt genomes with polymerase chain reaction (PCR), sequencing of PCR products, assembling of nucleotide sequences and annotation of the entire mt genomes.

### 2.1 Sequencing methods

(i) Isolation of mt genomes. Initial studies that investigated the mt genomes of living organisms obtained full sequences via the following steps: (1) isolation of mitochondria and the mtDNA within them using gradient centrifugation; (2) digestion of mtDNA with restriction enzymes; (3) ligation of digested fragments to a vector that was also cut with these restriction enzymes; (4) screening of recombinant clones containing target mtDNA; (5) sequencing of positive recombinants; and (6) assembling the mtDNA sequences. However, this approach is only suitable for living organisms with abundant materials resources or those that can be cultured *in vitro*. Specific methods have also been developed for the helminths. The mitochondria of *A. suum* adults from the intestines of pigs were isolated from the body wall muscle, or from mature eggs and lysed with 10% sarkosyl. mtDNA within the mitochondria was then isolated by cesium chloride-ethidium bromide centrifugation. Mitochondria of *C. elegans* worms (Bristol, N2 strain) cultured in a medium were also isolated as described for *A. suum*. DNA sequences were obtained from sets of deletion clones containing overlapping segments of the entire sequence of each complementary strand of the *A. suum* and *C. elegans* circular mtDNA molecules [11]. This method, however, is only suitable for sequencing helminth species that can be cultured *in vitro* or obtained in large quantities. Furthermore, these materials must be kept fresh to ensure that the entire mtDNA can be successfully isolated from the mitochondria [11,12]. It is also inappropriate to use this sequencing method if polymorphisms exist in the mtDNA from different individuals, such is often observed in nematodes. Such disadvantages have prevented mtDNA bio-markers from being widely applied to studies on helminthology, population genetics and other fields [13].

(ii) Library construction and selection of recombinant DNA. Mt genome sequencing for *Onchocerca volvulus* was achieved using this method. Steps included the extraction of total genomic DNA containing the mtDNA, followed by digestion with restriction enzymes, construction of the genomic DNA library with λEMBL4, screening of recombinant phage containing the entire mtDNA using a labeled partial cytochrome b (*cyt*b) gene PCR fragment, and primer-walking sequencing [14]. However, this approach is laborious and time-consuming thus, limiting its application [15].

(iii) Conventional PCR amplification. Despite the diverse uses mt genome sequencing offers, technical obstacles have restricted the full potential of mt markers. Typically, mt genome sequencing has relied on the isolation of mtDNA from the organism of interest. For large organisms, such as vertebrates, the purification of mtDNA in sufficient quantities to allow direct sequencing (without amplification and/or cloning) is achievable from fresh material. However, direct mt sequencing is not possible from specific tissues (e.g. neurons), biopsy samples, or material preserved for long periods, and cannot be applied to small invertebrates, such as parasitic worms.

Many methodologies for sequencing mt genomes have been based on the widely used PCR technique. These consist of the following main steps: PCR amplification of several fragments, which overlap, using pairs of primers designed according to conserved sequences, direct sequencing of PCR products or sequencing after cloning of fragments using primer-walking (Sanger method), and subsequent assembly of the entire mt genome sequence [16–18]. This approach avoids the need for the isolation and purification of mitochondria, which is laborious and time-consuming. Most genome sequences for parasitic worms are based on this approach. To improve the practical application of this method, some researchers developed specific modifications for various helminth species. For example, Tang et al. [19] successfully sequenced the complete mt genomes for *Thaumamermis cosgrovei* and *C. elegans* by using rolling circle amplification (RCA), which enables mtDNA to be

**Table 1** Helminth species (strains) with complete mitochondrial genomes characterized to date[a)]

| Name of species (strains) | Length (bp) | GenBank accession no. |
|---|---|---|
| Platyhelminthes (Phylum) | | |
| Trematoda (Class/16 species) | | |
| Digenea (Order) | | |
| *Clonorchis sinensis* Loose, 1907 | 13875 | NC_012894; FJ381664 |
| *Fasciola hepatica* Linnaeus, 1758 | 14462 | NC_002546; AF216697 |
| *Opisthorchis felineu*s Blanchard, 1895 | 14277 | NC_011127; EU921260 |
| *Paragonimus westermani* Braun, 1899 | 14965 | NC_002354; AF219379 |
| *Schistosoma haematobium* Bilarz, 1852 | 16901 | NC_008074; DQ157222 |
| *S. japonicum* Katsuurada, 1904 | 14085 | NC_002544; AF215860 |
| *S. mansoni* Sambon, 1907 | 14415 | NC_002545; AF216698 |
| *S. mekongi* Voge et al, 1978 | 14072 | NC_002529; AF217449 |
| *S. spindale* Montgomey, 1906 | 16901 | NC_008067; DQ157223 |
| *Trichobilharzia regenti* Horak et al, 1998 | 14838 | NC_009680; DQ859919 |
| Monogenea (Order) | | |
| *Benedenia hoshinai* Ogawa, 1984 | 13554 | NC_014591; EF055880 |
| *B. seriolae* Meserve, 1938 | 13498 | NC_014291; HM222526 |
| *Gyrodactylus derjavinoides* Malmberg et al, 2007 | 14741 | NC_010976; EU293891 |
| *G. salaries* Malmberg, 1957 | 14790 | NC_008815; DQ988931 |
| *G. thymalli* Zitnan, 1960 | 14788 | NC_009682; EF527269 |
| *Microcotyle sebastis* Coto, 1894 | 14,407 | NC_009055; DQ412044 |
| Cestoda (Class/28 species) | | |
| Cyclophyllidea (Order) | | |
| *Echinococcus granulosus* Batsch, 1786 (G1) | 13588 | NC_008075; AF297617 |
| *E. granulosus* (G4/horse strain) | 13598 | AF346403 |
| *E. granulosus* (G5/cattle strain) | 13717 | NC 011122; AB235846 |
| *E. granulosus* (G6/camel strain) | 13721 | NC_011121; AB208063 |
| *E. granulosus* (G7/pig strain) | 13719 | AB235847 |
| *E. granulosus* (G8/cervid strain) | 13717 | AB235848 |
| *E. multilocularis* Leuckart, 1863 | 13733 | NC_000928; AB018440 |
| *E. oligarthrus* Diesing, 1863 | 13791 | NC_009461; AB208545 |
| *E. shiquicus* Xiao et al, 2003 | 13807 | NC_009460; AB208064 |
| *E. vogeli* Rausch & Bernstein, 1972 | 13750 | NC_009462; AB208546 |
| *Hydatigera taeniaeformis* Bastch, 1786 (China) | 13647 | NC_014768; FJ597547 |
| Nematoda (Phylum) | | |
| Chromadorea (Class/38 species) | | |
| Ascaridida (Order) | | |
| *Anisakis simplex* Rudolphi, 1809 | 13916 | NC_007934; AY994157 |
| *Ascaris suum* Goeze, 1782 | 14284 | NC_001327; X54253 |
| *Toxocara canis* Stiles, 1905 | 14322 | NC_010690; AM411108 |
| *T. cati* Zeder, 1800 | 14092 | NC_010773; AM411622 |
| *T. malaysiensis* Gibbons, 2001 | 14266 | NC_010527; AM412316 |
| Diplogasterida (Order) | | |
| *Pristionchus pacificus* | 15954 | NC_015245; JF414117 |
| Oxyurida (Order) | | |
| *Enterobius vermicularis* Linnaeus, 1758 | 14010 | NC_011300; EU281143 |
| Rhabditida (Order) | | |
| *Ancylostoma caninum* Ercolani, 1859 | 13717 | NC_012309; FJ483518 |
| *A. duodenale* Creplin, 1845 | 13721 | NC_003415; AJ417718 |
| *Angiostrongylus cantonensis* Nomura, 1944 | 13497 | NC_013065 |
| *A. costaricensis* Morer & Cespedes, 1971 | 13585 | NC_013067 |
| *Bunostomum phlebotomum* Raillet,1902 | 13790 | NC_012308; FJ483517 |
| *Caenorhabditis briggsae* Dougherty, 1949 | 14420 | NC_009885; AC186293 |
| *C. elegans* Maupas, 1900 | 13794 | NC_001328; X54252 |
| *Chabertia ovina* Raillet & Henry, 1909 | 13682 | NC_013831; GQ888721 |
| *Cooperia oncophora* Ransom, 1907 | 13636 | NC_004806; AY265417 |
| *Cylicocyclus insignis* | 13828 | NC_013808; GQ888712 |
| *Haemonchus contortus* Cobb, 1898 | 14055 | NC_010383; EU346694 |
| *Heterorhabditis bacterophora* Poinar, 1976 | 18128 | NC_008534; EF043402 |
| *Mecistocirrus digitatus* Raillet, 1912 | 15221 | NC_013848; GQ888722 |
| *Metastronglyus pudendotectus* Wost, 1905 | 13778 | NC_013813; GQ888714 |
| *M. salmi* Gedoelst, 1923 | 1393 | NC_013815; GQ888715 |
| *Necator americanus* Stiles, 1903 | 13605 | NC_003416; AJ417719 |
| *Oesophagostomum dentatum* Molin, 1861 | 13869 | NC_013817; GQ888716 |
| *Steinernema carpocapsae* Woutus, 1982 | 13925 | NC_005941; AY591323 |
| *Strongyloides stercoralis* Bavay, 1876 | 13758 | NC_005143; AJ558163 |
| *Strongylus vulgaris* Looss, 1900 | 14301 | NC_013818; GQ888717 |

(*To be continued on the next page*)

(*Continued*)

| Name of species (strains) | Length (bp) | GenBank accession no. | Name of species (strains) | Length (bp) | GenBank accession no. |
|---|---|---|---|---|---|
| *H. taeniaeformis* (Germany ) | 13740 | unpublished* | *Syngamus trachea* von Siebold, 1836 | 14647 | NC_013821; GQ888718 |
| Multiceps multiceps (Leske, 1780) Hall, 1900 | 13693 | NC_012894;GQ228819* | *Teladorsagia circumcincta* Stadelman, 1894 | 14066 | NC_013827; GQ888720 |
| *Taenia asiatica* Eom & Rim, 1993 | 13703 | NC_004826; AF445798 | *Trichostrongylus axei* Raillet, 1909 | 13653 | NC_013824; GQ888719 |
| *T. crassiceps* Zedar, 1800 | 13503 | NC_002547; AF216699 | *T. vitrinus* Looss, 1905 | 13800 | NC_013807; GQ888711 |
| *T. hydatigena* Pallas, 1766 | 13492 | NC_012896;GQ228818* | Spirurida (Order) | | |
| *T. pisiformis* Bloch, 1780 | 13387 | NC_013844;GU569096* | *Brugia malayi* Buckloy, 1960 | 13657 | NC_004298; AF538716 |
| *T. saginata* Goeze, 1782 | 13670 | NC_009938; AY684274 | *Chandlerella quiscali* von Linstow, 1904 | 13757 | NC_014486; HM773029 |
| *T. solium* Linnaeus, 1758 | 13709 | NC_004022; AB086256 | *Dirofilaria immitis* Leidy, 1856 | 13814 | NC_005305; AJ537512 |
| *Hymenolepis diminuta* Rudolphi, 1819 | 13900 | NC_002767; AF314223 | *Onchocerca volvulus* Bickel, 1982 | 13747 | NC_001861; AF015193 |
| *Dipylidium caninum* Leuckart, 1863 | 13598 | unpublished* | *Steria digitata* Railliet et Henry, 1911 | 13839 | NC_ 014282; GU138699 |
| *Avitellina centripunctata* Rivolta, 1874 | 13559 | unpublished* | Tylenchida (Order) | | |
| *Thysaniezia giardi* Moniez, 1879 | 13768 | unpublished* | *Radopholus similes* Thorne, 1949 | 16791 | NC_013253; FN313571 |
| *Moniezia benedeni* Blanchard, 1891 | 13964 | unpublished* | Enoplia (Class/9 species) | | |
| *M. expansa* Blanchard, 1891 | 14133 | unpublished* | Dorylaimida (Order) | | |
| Pseudophyllidea (Order) | | | *Xiphinema americanum* Cobb, 1913 | 12626 | NC_005928; AY382608 |
| *Diphyllobothrium latum* Linnaeus, 1758 | 13608 | NC_008945; DQ985706 | Mermithida (Order) | | |
| *D. nihonkaiense* Yamane et al, 1986 | 13747 | NC_009463; AB268585 | *Agamermis* sp. BH-2006 Cobb et al, 1923 | 16561 | NC_008231; DQ665656 |
| *Spirometra erinaceieuropaei* Muller, 1937 | 13643 | NC_011037; AB374543 | *Hexamermis agrotis* Wang et al, 1986 | 24606 | NC_008828; EF368011 |
| Turbellaria (Class/1 species) | | | *Romanomermis culicivorax* Ross, 1976 | 26194 | NC_008640; EF154459 |
| Acoela (Order) | | | *R. iyengari* Welch, 1883 | 18919 | NC_008693; EF175764 |
| *Symsagittifera roscoffensis* Graff, 1891 | 14803 | NC_014578; HM237350 | *R. nielseni* Tsai & Grundmann, 1969 | 15546 | NC_008692; EF175763 |
| Acanthocephala (Phylum) | | | *Thaumamermis cosgrovei* Poinar, 1981 | 20013 | NC_008046; DQ520857 |
| Palaeacanthocephala (Class/1 species) | | | *Trelkovimermis spiculatus* Poinar, 1986 | 18030 | NC_008047; DQ520859 |
| Echinorhynchida (Order) | | | Trichocephalida (Order) | | |
| *Leptorhynchoides thecatus* Linton, 1891 | 13888 | NC_006892; AY562383 | *Trichinella spiralis* Railliet, 1895 | 16706 | NC_002681; AF293969 |

a) * represents mtDNA sequences completed by authors in this study.

amplified from a miniscule quantity of the initial sample before direct sequencing, analysis of restriction enzyme, gene cloning, and so on. Thus, this approach overcomes the limitations resulting from an insufficient amount of nematode materials [19]. Conventional PCR plays an important role in studies of helminth mt genomes, but it also has some disadvantages, such as sequencing the mt genome for a novel uncharacterized species without a reference sequence available, or failure to determine the new taxonomy status of a species because design of primers is difficult or complicated.

(iv) Long-range PCR (long-PCR) amplification. Mt genome sequencing usually relies on the isolation of high quality mtDNA from an organism, followed by sequencing (either directly or via PCR amplification) and sequence assembly to then determine the genome structure and gene order. For vertebrates and large invertebrates, relatively large amounts (micrograms or milligrams) of mtDNA can be isolated from individuals for subsequent PCR-coupled sequencing (either with or without cloning). For the much smaller helminths however, mt genome sequencing is often a considerable challenge because only a very small amount of DNA can be obtained from individual, substantial gene sequence heterogeneity often exists among individuals of a

species and the genomes are often AT-rich. Although an increased availability of material can be acquired through the "pooling" of multiple samples, this is often not possible due to the need to specifically examine one sample (e.g. from an individual nematode), the high rates of sequence polymorphisms occurring between or among multiple individuals (e.g. parasitic nematodes), or the unavailability of additional material. Therefore, to aid accurate and representative sequencing, PCR-coupled sequencing from DNA of individuals (without cloning) is preferable.

There is limited information on mt genomes and their organization for many metazoan organisms, particularly for the nematodes, which potentially represent four out of every five animals on the planet. Accumulated information on mt genome sequences depends largely on the availability of relatively simple and cost-effective techniques for sequencing. Recent advances in long-PCR amplification and the subsequent sequencing of mt amplicons by "primer-walking" have overcome the above-mentioned obstacles. This strategy is based on the long-PCR amplification of two overlapping fragments of the mt genome using two pairs of primers based on conserved regions and sequencing via primer-walking or cloning. This method also has the advantage of being applied to mtDNA sequencing for nematodes (usually 1–10 mm in length). For example, mt genome sequences were obtained of *Ancylostoma duodenale* and *Necator americanus* on the basis of the strategy [16]. Generally, the genomic DNA from a single nematode can be used five or more times for PCR, and PCR products can be used for direct sequencing, cloning and other analyses. As a result, only a single nematode is necessary for sequencing and other experiments, and sequencing is not restricted by the mtDNAs polymorphisms among individual nematodes [15,20,21]. Moreover, design of primers is relatively easy because only two pairs of primers are needed. For species whose mt genomes are well characterized, the process is rapid and relatively inexpensive. Therefore, this approach is likely to become a major method for the future sequencing of nematode mt genomes.

Nevertheless, for uncharacterized species (e.g. most parasites), *de novo* sequencing using "primer-walking" can be laborious, costly and inefficient. For highly AT-rich templates, such as the mt genomes of nematodes [6,9], primer-walking is significantly hampered by short sequence reads (~100 bp) and the limited availability of suitable regions for primer design. For most species, the depth of coverage (DOC) achieved using a standard bi-directional primer-walking approach is low (usually two times) [21,22] and limits opportunities for the detection of point mutations [13,20,22,23].

This approach is therefore not effective for high throughput sequencing, especially when *de novo* sequencing for an undetermined species or a species with a new taxonomic status is performed. The reason for this is that new primers need to be designed repeatedly during sequencing, requiring a much greater effort is required [21,22].

(v) High throughput sequencing. Advances in high-throughput sequencing technologies, particularly next-generation sequencing technology (NGS) such as massively parallel picoliter reactor sequencing (454 technology), offer a rapid and effective means of characterizing large numbers of mt genomes. These techniques are also suitable for overcoming many of the obstacles associated with primer-walking based methods. Recently, this sequencing approach was evaluated for the mt genomes of 12 socioeconomically important parasitic nematodes using small amounts of material from individual adult worms [21,22]. Complete mt genomes of each worm/species were amplified by long-PCR as two overlapping amplicons, then pooled and subsequently sequenced using a 454 Genome Sequencer FLX (Roche). The consensus mt genome sequences were each assembled automatically from thousands of individual "reads" based on a majority rule-threshold among all sequence data representing each contig.

Following sequencing of the complete mt genomes for *Haemonchus contortus*, *Anclystoma caninum* and *Cooperia oncophora*, the other 12 nematode full mt genomes were sequenced simultaneously (GenBank accession numbers GQ888711–GQ888722) [21,22]. The approach allows not only for *de novo* sequencing but also the multiplexing of tens or hundreds of mt genome templates in a single high throughput sequencing reaction. In addition, when coupled with whole genome amplification systems and/or laser micro-dissection technology, this approach might be applicable to assessing sequence variation or single nucleotide polymorphisms (SNPs) in mt genes among tissue types, providing insights into the heterogeneity within an individual and potentially increasing our understanding of mutation rates and inheritance. Furthermore, the extremely deep coverage obtained with NGS, which is a highly effective means to achieve previously undetermined mt genome sequences, allows for the detection of sequence heterogeneity between the multiple individuals or between the closely related species. This is particularly relevant for nematodes. Using NGS methodologies and scaffold, and *de novo* assembly strategies, Webb et al. [24] sought to comprehensively describe the extent and nature of divergence between the mt genomes of *Trichinella spiralis* and *Trichinella murrelli*. The former species represents an appreciable zoonotic risk owing to its capacity to establish persistent infections in domestic pigs, whereas the later is the most prevalent species in North American wildlife hosts and is of little risk to the safety of pork. The entire protein-coding region was sequenced, along with a portion of the highly repetitive non-coding region of the mitochondrial genome of *T. murrelli* with a combined average read depth of 250 reads. The accuracy of base calling, estimated from 13917 bp of coding region sequence was found to exceed 99.3%. NGS should therefore be considered a highly effective means to obtain previously unknown mitochondrial genome sequences [24].

### 2.2 Structure prediction and annotation of entire mt genome structure

Identifying coding regions of proteins, transfer RNA (tRNA) genes and rRNA genes can depend on the use of various specialized software combined with the BLAST and ClustalW programs. An open reading frame (ORF) of a protein can be found using ORF search software online (ORFinder from http://www.ncbi.nlm.nih.gov) or with DNAStar at a local computer (http://www.dnastar.com/). Prediction of secondary structure of tRNA can be carried out using special software such as tRNAscan-SE (http://mi.caspur.it/mitozoa/ tRNAscan-SE or http://lowelab.ucsc.edu/tRNAscan-SE/) and ARWEN (http://130.235.46.10/ ARWEN), which identifies most tRNAs from mt genomes with typical clover, D-loop or TV-loop morphology [25,26]. Any remaining tRNAs can be searched for using ClustalW alignment of their sequences combined with their folding structures. Furthermore, secondary structures of major non-coding regions can be predicated using RNAstructure and MFOLD software [27,28]. Finally, annotated sequence data can be imported into SEQUIN (available via http://www.ncbi.nlm.nih. gov/Sequin/) for final verification of the mt genomic structure, before subsequent direct submission to GenBank.

Additional online free annotated software, such as DOGMA, is available for parasitic worm mt genomes (http://dogma.ccbb.utexas.edu) [29]. Recently, an "in-house" prototypical bioinformatic system for the automated annotation and analysis of nematode mt genomic sequence datasets has also been produced, which will be greatly beneficial to future research [22].

## 3 Comparative mt genomics for helminths

Comparative helminth mt genomics mainly involves the assessment of differences and similarities in size, component, number and arrangement of genes, usage of codons, structure of tRNAs, among other criteria, between helminth species and also between helminths and other metazoans [30].

### 3.1 Comparison of mt genome size and composition

The mt genomes of worms are circular, similar to those of other eumetazoans. The complete mtDNA sequences of cestodes are 13.3–14.2 kb in length, highly compact and adenine and thymine (AT)-rich (comprising 70% of the mtDNA). Fluke mt genome sizes range from 13.9 to 16.9 kb with an AT-richness of 60%–70%, except for *P. westermani*, which comprises only 51.5% AT. The nucleotide compositions of the entire mtDNA sequences for helminths are biased toward T and A, with T being the most prevalent nucleotide and C the least common. The coding region comprises about 80%–95% of the full mt genome, of which, 70% belongs to the protein-coding region. Roundworms belonging to the Chromadorea have a 13.4–18.2 kb mt genome, while those from the Enoplia exhibit a 12.6–26.2 kb mt genome. The mtDNA of roundworms is also AT rich (68%–86%) [31,32].

### 3.2 Comparison of mt gene number and order

Each of the mt genomes sequenced for the platyhelminthes contains 36 genes lacking adenosine triphosphatase subunit 8 gene (*atp*8), including 12 protein-coding genes (*atp*6, the cytochrome *c* oxidase subunits 1, 2 and 3 [*cox*1–*cox*3], subunits of the nicotinamide dehydrogenase Q (60% of protein-coding region) including subunits 1–6 [*nad*1–*nad*6] and 4L [*nad*4L] and *cyt*b), 22 tRNA genes (two coding for leucine, and two coding for serine) and the small [*rrn*S] and large [*rrn*L] rRNA subunits. The mt gene arrangements of cestodes are nearly identical, whereas those of trematodes more variable. It is assumed that all genes are encoded in the heavy strand and transcribed in the same direction [33].

Nematodes also lack the *atp*8 gene that exists in higher animals with exception of *T. spiralis* [33]. Secernentean nematode mt genomes sequenced to date comprise 12 protein-encoded genes, two rRNA genes and 22 tRNA genes, which is identical to that of the platyhelminthes, whereas the mt genome of the Enoplia nematodes is distinctive. Most of Enoplia nematodes contain 12 protein-encoded, while others share 13 or 14 protein-coding genes with one to two having two copies. These nematodes also possess different numbers of tRNAs, ranging from 19 to 32. The genes of most roundworm mt genomes are located on the heavy (positive) chain, and transcribed and duplicated in a clockwise direction. Contrary to this, all genes in mt genome of *C. briggsae* are seated in a light (negative) strand [21,22,34]. In addition, genes encoding *nad*2, *nad*4, *nad*4L and *nad*5, and 10 tRNAs of *T. spiralis* are located on the negative chain, while nine protein-coding genes and other tRNA genes are found on the positive chain, which is similar to the coelomata metazoans. Other Enoplia nematodes share similar arrangements to that of *T. spiralis* mtDNA [22,33,35,36].

Although the gene arrangement within tapeworm mt genomes does not change, limited variation is occasionally observed in some worms. *trn*S2 of *Hymenolepis diminuta* and Anoplocephalids (unpublished data) is located in front of *trn*L1, which distinguishes it from other cestodes [37]. In addition, NR1 of *Taenia taeniaeformis* is situated between *trn*L1 and *trn*S2, while for other cestodes it is between *trn*Y and *trn*L1 (unpublished). With respect to the order or arrangement of genes within mt genomes, cestodes are closely related to the digenetic flukes such as *Fasciola hepatica*, *Paragonimus westermani*, *Clonorchis sinensis* and *Opisthorchis felineus*. They are also similar to *Gyrodactylus derjavinoides*, *Gyrodactylus salaris* and *Gyrodactylus thymalli* from the monogeneans. Although the number of genes in fluke mt genomes is consistent, their order is variable when

compared with that of the cestodes, especially the diverse and complex arrangement for *Schistosoma haematobium*, *Schistosoma mansoni* and *Schistosoma spindale* [38–44].

mtDNA gene size within the roundworms is relatively fixed and conserved. However, their number and arrangement varies largely, which is significantly different from that of cestodes and trematodes. Nematodes are grouped into 11 different gene arrangements (GA1–GA11) (Figure 1) [45]. Approximate 50% of species share the GA6 gene order, including *Setaria digitata*, *O. volvulus*, *Dirofilaria immitis* and *Brugia malayi.* This gene arrangement is distinctly different from the other nematodes whose mt genomes have been sequenced thus far. Nematodes from the Enpolia have highly diverse numbers and arrangements in mt genes [34,36].

### 3.3 Comparison of mt gene codon usage

Mitochondrial protein-coding regions of flatworms and roundworms are based on translation tables 9 and 5 of the genetic codons, respectively. Within table 9, ATG and GTG are the most common initiation codons while TAA and TAG are the most dominant termination codons. In addition to these two start codons, an unusual start codon GTT has been suggested for *cox*1 of *Echinococcus granulosus* (G4 strain) [46] and *H. diminuta* [37], and *atp*6 of *T. pisiformis*. Moreover, AGR encodes for serine, AAA for asparagine, and TGA for tryptophan. ATT, ATC, ATA and TTG can be used as alternative initiation codons besides ATG and GTG, while AGR codes for serine, TGA for tryptophan, and ATA for methionine instead of isoleucine. In addition, the initiation codon TTT is unique to the *S. digitata* mt genome and its four protein-coding genes use this codon as a translation initiation codon [45]. Abbreviated stop codons are also found in the mt protein-coding genes of helminths, for example, TAA for *nad*3 of *T. pisiformis* may be completed by the addition of 3' A residue to the mRNA post transcription [47,48].

The protein-coding genes of helminth mt genomes are biased toward using amino acids encoded by codons rich in T, A and G. The T-rich codons (with≥2 Ts in a triplet), comprising Phe (TTT and TTC), Leu (TTA, TTG and CTT), Ile (ATT), Val (GTT), Tyr (TAT), Ser (TCT) and Cys (TGT), account for approximately half of the total amino acid composition. A-rich and G-rich codons (with≥2 As and Gs, respectively) represent relatively abundant amount of the total amino acid composition. In contrast, the proportion of C-rich codons (with≥2 Cs) is much lower [47].

### 3.4 Comparison of mtRNA secondary structure

Mitochondrial tRNAs of flatworms are approximately 53–76 nucleotides in length and form two different structures. One is the typical clover shape, which is observed for 18 tRNAs, and the other is a D-loop structure lacking a paired DHU arm for *trn*C, *trn*S1, *trn*S2 and *trn*R. *trn*C from *H. diminuta* and *trn*R from *Paragonimus westermani* have a clover structure [31,37,38], while the mt *trn*S1 and *trn*S2 from *F. hepatica* and *C. sinensis* are D-loop and/or clover form [31,38]. *trn*Fs from most flukes have a clover structure but those from *Schistosoma mekongi* and *Schistosoma malayensis* form a TV-loop, lacking variable arms and a paired stem [31,33,41].

In contrast to the flatworms, the mtRNAs for the roundworms are relatively short (52–65 nucleotides) and form three kinds of secondary structures. Most tRNAs form a TV-loop, which is unique in the animal kingdom, and *trn*S1 and *trn*S2 form a D-loop. In addition, eight tRNAs (*trn*R, D, K, I, L1, L2, M and W) from *T. spiralis* have a clover structure [33,36,49].

Anticodons in 22 tRNAs for all published mt genomes of helminths are very conserved and lack mutations. For the tRNAs of the taeniid cestodes, the first two nucleotides in front of the anticodons are almost always thymidines, and the first nucleotide following the anticodons is typically A or G, while the second nucleotide is typically G [47].

### 3.5 Comparison of the main non-coding regions and their secondary structures

An overwhelming majority of helminth mt genomes contain

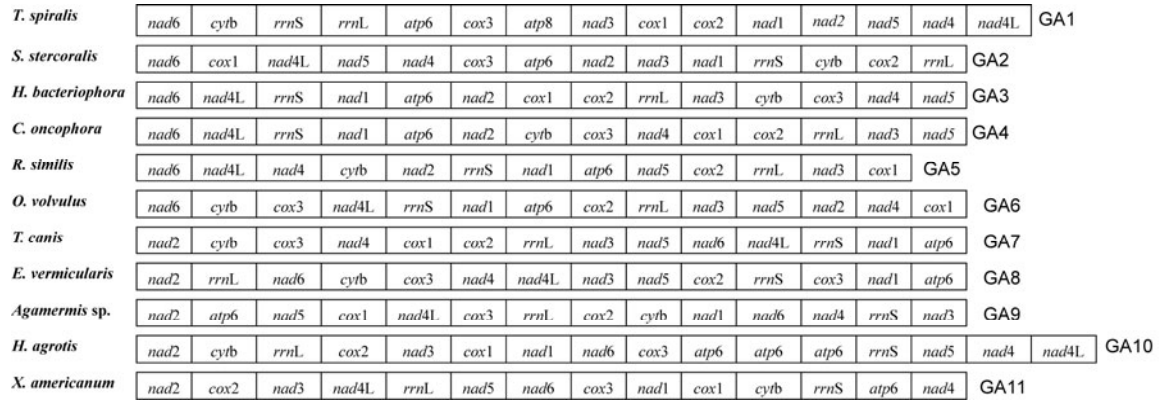


**Figure 1** Comparison of mtDNA protein-encoding and ribosome RNA gene (GA) arrangement for nematodes [45].

two main non-coding regions (NR), which are different in size and sequence. These two regions are usually AT-rich and contain more than one inverted sequence, which can be folded into a complex and stable stem-loop structure. NRs might also control the origin and duplication of mtDNA [11,17,37]. One major NR in *Trichobilharzia regenti* contains three 184 bp repeat sequences, each of which can form into a stable stem-loop structure [42]. Certain tapeworm mt genomes, such as *Moniezia expansa,* possess 15 variable tandem repeat sequences (ATTATGATGTATAATAGGTATAGTGTATTAA). Within the NR1 of the *E. multilocularis* mt genome, an AATTTATCCGGTTTGATGTGCCT unit and its reverse unit AGGCACATCAAACCGGATAAATT both exist as two repeats. Both units are positioned together during interphase, separated by other short sequences [17]. These stem-loop structures are similar to those of the conserved CSB1–CSB3 regions close to the D-region of vertebrate mtDNA, which can be used as binding sites for DNA and protein (transcribed initiation factor). Therefore, the AT-rich NRs are potentially the regions that control origin and duplication within helminth mt genomes [17].

### 3.6 Comparison of the size overlap between *nad*4L and *nad*4 genes

The overlap between mt *nad*4L and *nad*4 genes is consistently 40 bp for most cestodes except for the genus *Taenia*, which have an overlap of 34 bp [47]. For trematodes, the length of the overlap is mainly 28 or 40 bp. Exceptions include *S. mekongi* (64 bp), *S. japonicum* and *T. regenti* (both 37 bp) [31]. In contrast, the *nad*4L and *nad*4 genes for most nematodes are not positioned side by side, or *nad*4 is in located in front of *nad*4L, as is the case for *T. spiralis* and *Hexamermis agrotis* (NC_008828) [33].

## 4 Helminth mt genome applications

Mitochondrial genome sequences have widely been applied to various areas of helminth research, including species identification, molecular taxonomy, evolutionary history and relationships, and diagnosis of parasitic diseases.

### 4.1 Molecular taxonomy and identification of helminth species

The classification of tapeworms belonging to the family Taeniidae has previously been controversial because of the paucity of adult phenotypic characters and the morphological plasticity of larvae that develop within various intermediate hosts. Recent advances in DNA amplification and sequencing have allowed for the development of molecular-based taxonomic classification. Database catalogs of taxonomically useful DNA sequences known as DNA barcoding have facilitated non-morphological identification of parasite species.

The beef tapeworm *T. saginata* is of medical and veterinary importance because it causes bovine cysticercosis and human taeniasis. *Taenia asiatica* is the most recently described *Taenia* tapeworm, whose classification is based on morphological characteristics and has a life cycle that differs from that of *T. saginata*. Molecular biological data and DNA sequence data have revealed that *T. asiatica* is closely related to *T. saginata*, but confirmed that it is a distinct species. The overall sequence difference in the full mt genomes between *T. saginata* and *T. asiatica* is 4.6%, while they differ by 11% with *Taenia solium*. Such mt gene sequence divergence is frequently used as a "yardstick" for closely related species. The criterion that most of congeneric species show greater than 2% divergence supports the species validity of *T. asiatica* [47,50–52]. Analysis of the full-length 18S rDNA sequences of *T. asiatica* and *T. saginata* revealed a genetic difference of only 0.7%, making it difficult to distinguish these two species on the basis of this gene. As a result, the Asian *Taenia* (now known as *T. asiatica*) used to be considered as the same species as *T. saginata* or as a subspecies or strain (*T. saginata asiatica*) [53].

Molecular analyses using the mtDNA sequences of the *cox*1 and *nad*1 genes showed that *E. granulosus* could be divided into ten main genotypes (G1–G10), as corresponding to the definition of a strain [54–59]. Recent taxonomic revisions indicated that *E. granulosus* is an oversimplified species in which four or five cryptic species were intermixed [60]. Sequencing of the complete mt genomes has further facilitated phylogenetic studies on *E. granulosus*. Comparative analysis of *E. granulosus* (sheep, horse, camel, lion and pig strains) mt genomes has led to recent taxonomic revision. Greater than 10% mtDNA divergence has been reported to exist between *Echinococcus equinus* (G4) and other *E. granulosus* genotypes. While more than a 6% nucleotide difference was observed for *Echinococcus ortleppi* (G5) and other *E. granulosus* strains not including *E. equinus*. *Echinococcus canadensis,* including G6–G8 and G10, were found to be significantly from *E. granulosus* s. s. G1 with a divergence of more than 10%. By reviving synonym and subspecies names, this species has now been split into *E. granulosus* s. s. (genotypes G1–G3), *E. equines* (G4), *E. ortleppi/E. felidis* (G5) and *E. canadensis* (G6–G10) [55,61–64]. *E. granulosus* s. s. is distantly related to *E. equinus,* and *E. ortleppi* and the genotypes G6, G7 and G8 form a monophyletic group. In particular, the close relationship of G6, G7 and G8 suggests that these genotypes (camel, pig and cervid strains, respectively) should be unified into *E. canadensis*. Furthermore, Hüttner et al. [65] reported that *E. granulosus* s. s. and *Echinococcus felidis* are sister species (Figure 2).

Debates continue over the extent to which the parasitic trichostrongylids *Haemonchus placei* and *H. contortus* hybridize in nature, and whether they each deserve species status. Mitochondrial *nad*4 gene sequences from individuals

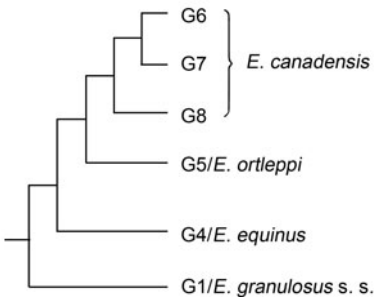


**Figure 2** Taxonomic status and evolutionary relationships of *E. granulosus* intraspecies based on full-length sequences of their mt genomes [65].

of each putative species collected from populations around the United States indicated that the two species are distinctive species at the mtDNA level, with 16% nucleotide divergence. Furthermore, there was no evidence of introgressive hybridization occurring in wild populations [66].

### 4.2 Species origin and evolution

The relationships between the three classes of Neodermata (parasitic Platyhelminthes) are much debated and restrict our understanding of the evolution of parasitism and contingent adaptations. The historic view of a sister relationship between Cestoda and Monogenea (Cercomeromorphae; larvae bearing posterior hooks) has been dismissed and continued evidence has been provided against the existence of monogenean monophyly. Nucleotide analyses of the complete mt genomes for *Benedenia seriolae* (Monogenea: Monopisthocotylea: Capsalidae), the first complete non-gyrodactylid monopisthocotylean mt genome to be sequenced, the nucleotide sequence for some of the mt protein coding genes for a second capsalid, *Neobenedenia* sp., and *M. sebastis* (Monogenea; Polyopisthocotylea; Microcotylidae), along with all available platyhelminth mt genomes provided new phylogenetic hypotheses that have strongly influenced our perspectives on the evolution of diet in the Neodermata. The analyses do not support monogenean monophyly but confirm that the Digenea and Cestoda are each monophyletic and sister groups. This improves our understanding of the relationships between these species and fundamentally enhances our knowledge of the evolution of parasitism in the Neodermata and in particular, the evolution of diet (the gene arrangement of flatworm mt genomes is shown in Figure 3) [39,67,68].

Traditional *Schistosoma* species phylogenies have been characterized mainly on the basis of egg morphology, intermediate host specificity and biogeography, and comprise four groups represented by the species *S. japonicum* (including *S. malayensis*, *S. sinensium*, *S. mekongi*, and *S. ovuncatum*), *S. mansoni* (including *S. rodhaini*, *S. hippopotami*, and *S. edwardiense*), *S. indicum* (including *S. incognitum*, *S. nasale* and *S. spindale*) and *S. haematobium* (such as *S. bovis* and *S. margrebowiei*). Analyses of the mt genomes of *S. japonicum*, *S. mansoni*, *S. spindale* and *S. haematobium* respectively from each group show that the mt genomes of *S. mansoni*, *S. spindale* and *S. haematobium* possess a similar gene order. These species are all distributed in West and South Asia and Africa and have significantly different gene arrangement from that of members of the *S. japonicum* species group. These data support the "out of East and Southeast Asia" hypothesis for the origin of this genus. The unique order of protein-encoding genes also further differentiates the majority of *Schistosoma* species from other Digenea and Cestoda that have been fully characterized to date (Figure 3) [31].

### 4.3 Molecular epidemiology, geography and ecology

DNA-based approaches for the identification of parasites have also been implemented within the fields of molecular epidemiology and ecology. In ecological studies on *Echinococcus* spp., genetic markers are still needed to trace the spatial spread of these parasites in synanthropic and nonsynanthropic habitats. A commonly applied marker for the assessment of population genetic structures is the highly polymorphic microsatellite, consisting of short tandemly repeated DNA. Single locus microsatellite markers are especially important in analyzing the genetic relationships between parasites. Although some single locus markers have been isolated from *E. multilocularis* and *E. granulosus*, their total number is insufficient for ecological use. A haploid maternally inherited mtDNA marker has been alternatively used for population genetic studies. Phylogeographical analysis using mtDNA markers demonstrated that *E. multilocularis* populations are generally divided into European, Asian and North American clades, which were likely caused by the vicariance of host foxes during the Pleistocene ice ages [69]. Moreover, the statistical parsimony network of mtDNA haplotypes illustrated that founder effects had recently occurred in the Chinese populations of *E. granulosus* s. s. and *E. multilocularis*, and that *E. shiquicus*, which requires the Tibetan fox and plateau pika as the definitive and intermediate host, respectively, was completely different from the former two species [70,71]. In different geographical areas and hosts, *E. granulosus* may have distinctive hereditary traits. These strains or even species have differences in larval morphology, pathogenicity to humans, range of hosts and epidemiology, which are important to consider when developing vaccines, diagnostic reagents and therapies (summarized in Table 2) [72]. In China, two major genotypes are often identified, including G1 (sheep strain), which uses sheep, humans and cattle as the intermediate host, and G6 (camel strain), which uses humans, cattle and camels as the intermediate host [60,66,72].

Minimal genetic variability was initially found within *T. solium* populations on the basis of the short nucleotide sequences of the mt *cox*1 and a nuclear gene that was developed as the diagnostic antigen Ts14 [73]. However, another

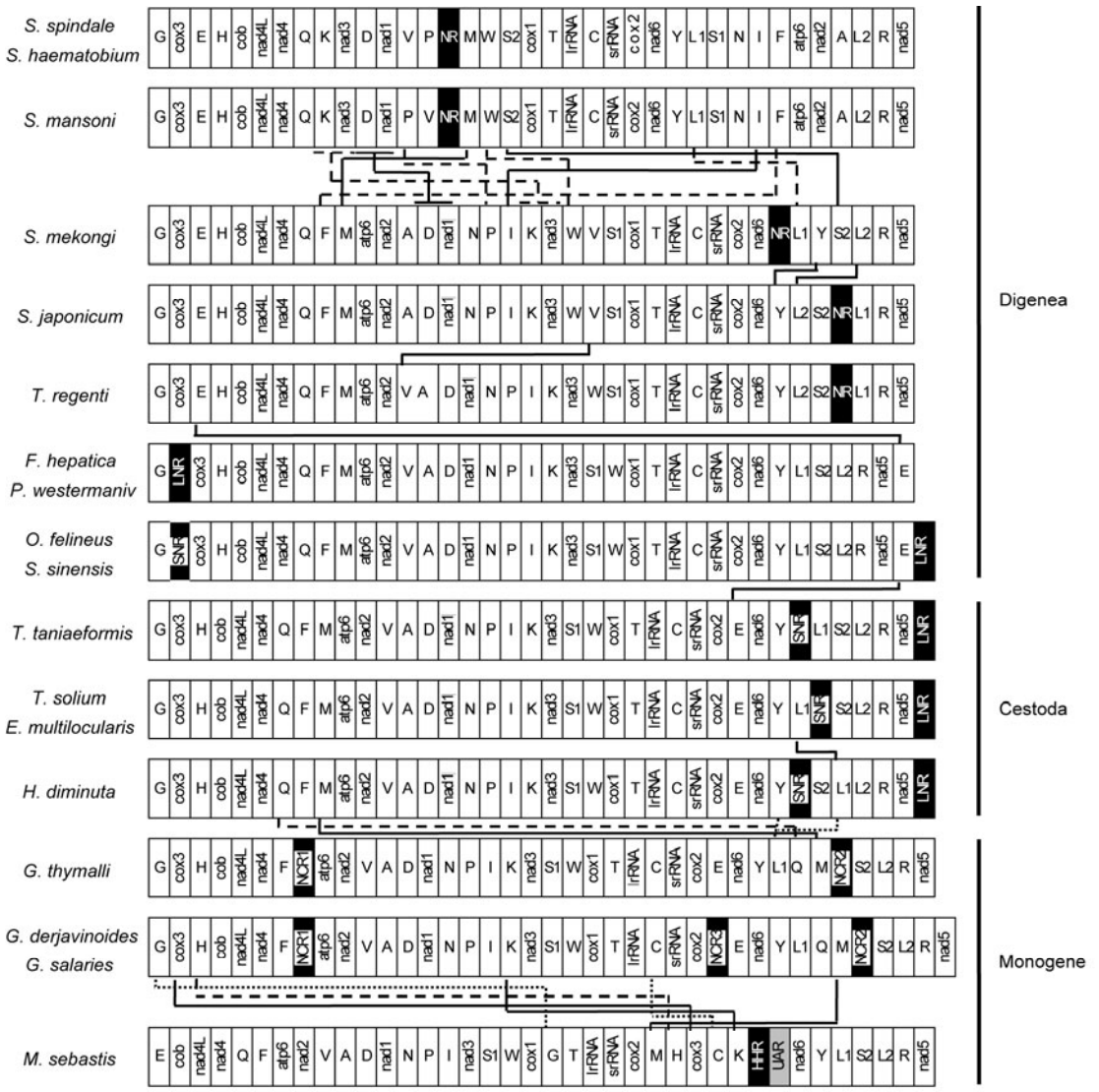


**Figure 3** Comparison of mt gene arrangements for representative flatworms (cestodes and trematodes), and their genetic relationships [39,67]. Black or gray box stands for major noncoding regions in mt genomes.

survey using the complete nucleotide sequences of mt *cox*1 and *cyt*b genes revealed that *T. solium* is divided into two main geographic clades [74]. One is widely distributed in Latin America and Africa, while the other is restricted to Asia. Such a geographical pattern suggests that *T. solium* was recently introduced into Latin America and Africa from Europe during the colonial age, which started 500 years ago, and that this tapeworm independently spread into Asia. Recently, this hypothesis was further supported by a new analysis of published data [75].

### 4.4 Diagnosis of parasitic diseases

Cryptic species, which are morphologically similar yet genetically different, may not be easily distinguishable based on conventional morphological methods. In addition, morphological characteristics of a species are often influenced by its habitat within the host(s), host species, and environmental factors, among others. Therefore, selection of suitable and reliable molecular markers is one of the most effective ways of identifying parasitic worms. Mitochondrial genes should be considered as useful molecular markers for the investigation of molecular epidemiology and diagnosis of helminths.

Significant differences in the life histories and mt genome sequences of the human hookworms *A. duodenale* and *N. americanus* necessitate their differentiation in epidemiological studies and the design of control programs. Current methods of characterization require time-consuming, labor-intensive techniques based on their life histories. However, Zhan et al. [76] conveniently PCR amplified 585-bp fragments of *cox*1 gene from individual hookworm eggs, larvae, and adults, suggesting that this technique is useful for identifying mixed infections containing equal

**Table 2** Strains (genotypes) of *E. granulosus* and their geographic distribution [72][a)]

| Strains (genotypes, G) | Intermediate hosts and aberrant hosts | Definitive hosts | Probable geographic distribution |
|---|---|---|---|
| Sheep strains (G1) | Sheep, cattle, pig, camel, goat, macropod, human | Dog, fox, dingo, jackal, hyena | Australian mainland, Europe, USA, New Zealand, Africa, China, Middle East, South America, Russian Federation |
| Tasmanian sheep strain (G2) | Sheep, cattle?, human? | Dog, fox | Tasmania, Argentina |
| Buffalo strain (G3) | Buffalo, cattle?, human | Dog, fox? | Asia |
| Horse strain (G4) | Horses and other equines | Dog | Europe, Middle East, South Africa, New Zealand?, USA? |
| Cattle strain (B5) | Cattle, human | Dog | Europe, South Africa, India, Sri Lanka, Russian Federation |
| Camel strain (G6) | Camel, goat, cattle?, human? | Dog | Middle East, Africa, China, Argentina, Canada |
| Pig strain (G7) | Pig, human? | Dog | Europe, Russian Federation, South Africa, Canada |
| Cervid strain (G8) | Elk, deer, reindeer, human? | Wolf, dog | South America, Eurasia, Canada |
| Pig strain (G9) | Pig, human? | Dog? | Poland, Canada |
| Cervid strain (G10) | Cervid, human | Wolf, dog | Canada |
| Lion strain* | Zebra, wildebeest, warthog, bushpig, buffalo, various antelopes, giraffe?, Hippopotamus? | Lion | Africa |

a)?, unclear; *, no detailed genetic characterization.

amounts of eggs from each species. Furthermore, this technique is rapid, technically simple and sensitive, and allows for the accurate identification of human hookworms in epidemiological field studies.

### 4.5 Control strategy for parasitic diseases

*T. solium*, *T. saginata* and *T. asiatica* are known causative agents of taeniasis in humans. *T. solium* also causes cysticercosis in humans. Neurocysticercosis caused by larval *T. solium* cysticerci developed in the central nervous system is the most serious of these diseases and is characterized by diverse neurologic symptoms, which commonly includes epileptic seizure. Furthermore, gravid proglottids filled with eggs expelled from tapeworm carriers serve as a source of infection for intermediate hosts, particularly in developing countries where sanitary conditions are poor. Finally, these three tapeworms occur sympatrically in the same area [77–79] and therefore, early and accurate detection and adequate treatment of taeniasis is important for the prevention of cysticercosis infections [80]. Reliable epidemiological information is also important for effective control of taeniasis or cysticercosis, including the use of accurate tools for parasite identification, which currently need to be developed.

Most recently, a new method based on the thymine bases of mitochondrial genes has been developed for the comprehensive differential diagnosis of *T. saginata*, *T. asiatica*, and two genotypes of *T. solium* [74]. In addition, a simple and reliable multiplex PCR has been established for differential diagnosis of the causative agents of taeniasis and cysticercosis [81]. Yamasaki et al. [82] developed and evaluated a loop-mediated isothermal amplification (LAMP) assay based on a nuclear cathepsin L-like cysteine peptidase (*clp*) gene and mt *cox*1 gene for differentiation and rapid diagnosis of *Taenia* species infections. LAMP detected a single copy of the target gene or five eggs of *T. asiatica* and *T. saginata* per gram of feces, similar to the sensitivity achieved using conventional PCR methods. Due to their rapid, simple, specific, and sensitive detection of *Taenia* species, LAMP assays are valuable tools, which might be easily applicable for the control and prevention of taeniasis and cysticercosis in countries where these diseases are endemic [81–86].

## 5 Future studies on helminth mt genomes

Increasing numbers of helminth mt genome sequences are likely to be characterized in the near future. Since the first entire helminth mt genome was described, very few complete mt genomes from worms were sequenced during the following 20 years. However, more than 30 complete mtDNA sequences, which account for one third of all characterized mtDNAs for worms, were only recently sequenced between 2009 and 2010. This means that advances in sequencing of entire mtDNAs have rapidly increased. Nevertheless, the number of worm species with full mt genomes characterized so far is a very small proportion of all of the described helminth species. It is therefore likely that future research will focus on rapidly increasing in the amount helminth mt genome sequence data available.

Next generation methodologies associated with high throughput sequencing are also likely to increase in use as these techniques mature and their cost decreases. In addition,

DNA-based technology has improved substantially, allowing for very small DNA quantities from worms to be sufficient for high throughput sequencing of mt genomes. Furthermore, rapid development within the field of bioinformatics will further enhance the analysis of next generation sequencing data for the characterization of entire mt genomes.

Although entire mt genomes are relatively small compared with nuclear genomes, the high-throughput and custom built bioinformatic platform that has been established represents an efficient method that can be applied to any group of organisms. Computer software for automated annotation will also be further developed and employed. The development of new and special software for analyses and automated annotation of cestodes, trematodes and nematodes suitable for different purposes are also needed for future mt genomic investigations.

Mitochondrial genome data already deposited in databases further needs to be mined. Because mt genome sequences have become important DNA barcoding or biomolecular markers for living organisms, there has been a rapid expansion in data available in genetic databases, including those specific for mt genomes. Available data and specific databases are likely to expand greatly in the future.

Mitochondrial genes and genomes are useful as molecular markers for systematic and population genetic studies across a broad range of worm taxa. Among the pathogenic worms, mt genetic data has been utilized in many epidemiological studies, often allowing for the investigation of pathogen evolution (e.g. population or "strain" differentiation and speciation). In addition, the characteristics of diseases (e.g. pathogenesis, host affiliations, virulence and drug resistance), potential links to disorders, and other aspects related to helminth mt genomes need to be further investigated in future.

## 6 Conclusions

Mitochondrial genome sequences can display a number of unique features, which have been used for studying the fundamental aspects of genetics and genomics. As a result, exploring mt genes and genomics of worms has attracted much research attention. Increased numbers of complete mt genome sequences have now been characterized for many helminth species. Understanding helminth mt genomics is crucial for improving our knowledge of molecular taxonomy and identification, population genetics, evolutionary relationships, epidemiology, and diagnosis of parasitic worms, thereby providing insights into the relationships of closely related and diverged species, and also novel control strategies for parasitic diseases.

*This work was supported by the Governmental Public Industry (Agriculture) Research Special Funds for Projects (200903036-07), the National Key Project of Scientific and Technical Supporting Programs (2007BAD40B04), and the National Key Laboratory Program (SKLVEB2008ZZKT). The authors thank Professor D. Timothy J. Littlewood, Department of Zoology, The Natural History Museum, for critical comments and suggestions on this paper.*

1 Keiser J, Utzinger J. Food-borne trematodiases. Clin Microbiol Rev, 2009, 22: 466–483
2 Pawlowski Z S. Cestodiases: Taeniasis, cysticercosis, diphyllobothriasis, hymenolepiasis, and others. In: Warren K S, Mahmoud A A F, eds. Tropical and Geographical Medicine. 2nd ed. New York: McGraw-Hill Information Services, 1990
3 Chai J Y, Murrell K D, Lymbery A J. Fish-borne parasitic zoonoses: Status and issues. Int J Parasitol, 2005, 35: 1233–1254
4 Koenning S R, Overstreet C, Noling J W, et al. Survey of crop losses in response to phytoparastic nematodes in the United States for 1994. J Nematol, 1999, 31: 587–618
5 Olson P D, Tkach V V. Advances and trends in the molecular systematic of the parasitic platyhelminthes. Adv Parasitol, 2005, 60: 165–243
6 Hu M, Gasser R B. Mitochondrial genomes of parasitic nematodes—Progress and perspectives. Trends Parasitol, 2006, 22: 78–84
7 Boore J L, Brown W M. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev, 1998, 8: 668–674
8 Boore J L. Animal mitochondrial genomes. Nucleic Acids Res, 1999, 27: 1767–1780
9 Hu M, Chilton N B, Gasser R B. The mitochondrial genomics of parasitic nematodes of socio-economic importance: Recent progress, and implications for population genetics and systematics. Adv Parasitol, 2003, 56: 133–212
10 Gasser R B, Newton S E. Genomic and genetic research on bursate nematodes: Significance, implications and prospects. Int J Parasitol, 2000, 30: 509–534
11 Okimoto R, Macfarlane J L, Wolstenholme D R. Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*. Nucleic Acids Res, 1990, 18: 6113–6118
12 Okimoto R, Macfarlane J L, Clary D O, et al. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics, 1992, 130: 471–498
13 Littlewood D T, Gasser R B. Toward next-generation sequencing of mitochondrial genomes—Focus on parasitic worms of animals and biotechnological implications. Biotechnol Adv, 2010, 18: 151–159
14 Keddie E M, Higazi T, Unnasch T R. The mitochondrial genome of *Onchocerca volvulus*: Sequence, structure and phylogenetic analysis. Mol Biochem Parasitol, 1998, 95: 111–127
15 Burger G, Lavrov D V, Forget L, et al. Sequencing complete mitochondrial and plastid genomes. Nat Protoc, 2007, 2: 603–614
16 Hu M, Chilton N B, Gasser R B. The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). Int J Parasitol, 2002, 32: 145–158
17 Nakao M, Yokoyama N, Sako Y, et al. The complete mitochondrial DNA sequence of the cestode *Echinococcus multilocularis* (Cyclophyllidea: Taeniidae). Mitochondrion, 2002, 1: 497–509
18 Le T H, Blair D, Agatsuma T, et al. Phylogenies inferred from mitochondrial gene orders–A cautionary tale from the parasitic flatworms. Mol Biol Evol, 2000, 17: 1123–1125
19 Tang S, Hyman B C. Rolling circle amplification of complete nematode mitochondrial genomes. J Nematol, 2005, 37: 236–241
20 Hu M, Jex A R, Campbell B E, et al. Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. Nat Protoc, 2007, 2: 2339–2344
21 Jex A R, Hu M, Littlewood D T, et al. Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genomics, 2008, 9: 11
22 Jex A R, Hall R S, Gasser R B, et al. An integrated pipeline for

next-generation sequencing and annotation of mitochondrial genomes. Nucleic Acids Res, 2010, 38: 522–533
23 Hu M, Chilton N B, Gasser R B. Long PCR-based amplification of the entire mitochondrial genome from single parasitic nematodes. Mol Cell Probes, 2002, 16: 261–267
24 Webb K M, Rosenthal B M. Next generation sequencing of the *Trichinella murrelli* mitochondrial genome allows comprehensive comparison of its divergence from the principal agent of human trichinellosis, *Trichimella spiralis*. Infect Genet Evol, 2011, 11: 116–123
25 Lowe T M, Eddy S R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res, 1997, 25: 955–964
26 Laslett D, Canback B. ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics, 2008, 24: 172–175
27 Mathews D H. Predicting a set of minimal free energy RNA secondary structures common to two sequences. Bioinformatics, 2005, 21: 2246–2253
28 Mathews D H, Disney M D, Childs J L, et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA, 2004, 101: 7287–7292
29 Wyman S K, Jansen R K, Boore J L. Automatic annotation of organellar genomes with DOGMA. Bioinformatics, 2004, 20: 3252–3255
30 Lupi R, de Meo P D, Picardi E, et al. MitoZoa: A curated mitochondrial genome database of metazoans for comparative genomics studies. Mitochondrion, 2010, 10: 192–199
31 Littlewood D T, Lockyer A E, Webster B L, et al. The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. Mol Phylogenet Evol, 2006, 39: 452–467
32 Steinauer M L, Nickol B B, Broughton R, et al. First sequenced mitochondrial genome from the phylum Acanthocephala (*Leptorhynchoides thecatus*) and its phylogenetic position within Metazoa. J Mol Evol, 2005, 60: 706–715
33 Lavrov D V, Brown W M. *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, 2001, 157: 621–637
34 He Y, Jones J, Armstrong M, et al. The mitochondrial genome of *Xiphinema americanum* sensu stricto (Nematoda: Enoplea): Considerable economization in the length and structural features of encoded genes. J Mol Evol, 2005, 61: 819–833
35 Jex A R, Waeschenbac A, Hu M, et al. The mitochondrial genomes of *Ancylostoma caninum* and *Bunostomum phlebotomum*—Two hookworms of animal health and zoonotic importance. BMC Genomics, 2009, 10: 79
36 Powers T O, Harris T S, Hyman B C. Mitochondrial DNA sequence divergence among *Meloidogyne incognita*, *Romanomermis culicivorax*, *Ascaris suum*, and *Caenorhabditis elegan*s. J Nematol, 1993, 25: 564–572
37 von Nickisch-Rosenegk M, Brown W M, Boore J L. Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: Gene arrangements indicate that Platyhelminths are Eutrochozoans. Mol Biol Evol, 2001, 18: 721–730
38 Shekhovtsov S V, Katokhin A V, Kolchanov N A, et al. The complete mitochondrial genomes of the liver flukes *Opisthorchis felineus* and *Clonorchis sinensis* (Trematoda). Parasitol Int, 2010, 59: 100–103
39 Perkins E M, Donnellan S C, Bertozzi T, et al. Closing the mitochondrial circle on paraphyly of the Monogenea (Platyhelminthes) infers evolution in the diet of parasitic flatworms. Int J Parasitol, 2010, 40: 1237–1245
40 Plaisance L, Huyse T, Littlewood D T, et al. The complete mitochondrial DNA sequence of the monogenean *Gyrodactylus thymalli* (Platyhelminthes: Monogenea), a parasite of grayling (*Thymallus thymallus*). Mol Biochem Parasitol, 2007, 154: 190–194
41 Huyse T, Plaisance L, Webster B L, et al. The mitochondrial genome of *Gyrodactylus salaris* (Platyhelminthes: Monogenea), a pathogen of *Atlantic salmon* (Salmo salar). Parasitology, 2007, 134: 739–747
42 Webster B L, Rudolfova J, Horak P, et al. The complete mitochondrial genome of the bird schistosome *Trichobilharzia regenti* (Platyhelminthes: Digenea), causative agent of cercarial dermatitis. J Parasitol, 2007, 93: 553–561
43 Huyse T, Buchmann K, Littlewood D T. The mitochondrial genome of *Gyrodactylus derjavinoides* (Platyhelminthes: Monogenea) —A mitogenomic approach for Gyrodactylus species and strain identification. Gene, 2008, 417: 27–34
44 Le T H, Humair P F, Blair D, et al. Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. Mol Biochem Parasitol, 2001, 117: 61–71
45 Yatawara L, Wickramasinghe S, Rajapakse R P, et al. The complete mitochondrial genome of *Setaria digitata* (Nematoda: Filarioidea): Mitochondrial gene content, arrangement and composition compared with other nematodes. Mol Biochem Parasitol, 2010, 173: 32–38
46 Van der Veer M, de Vries E. A single nucleotide polymorphism map of the mitochondrial genome of the parasitic nematode *Cooperia oncophora*. Parasitology, 2004, 128: 421–431
47 Jia W Z, Yan H B, Guo A J, et al. Complete mitochondrial genomes of *Taenia multiceps*, *T. hydatigena* and *T. pisiformis*: Additional molecular markers for a tapeworm genus of human and animal health significance. BMC Genomics, 2010, 11: 447
48 Nakao M, Sako Y, Ito A. The mitochondrial genome of the tapeworm *Taenia solium*: A finding of the abbreviated stop codon U. J Parasitol, 2003, 89: 633–635
49 Wolstenholme D R, Okimoto R, Macfarlane J L. Nucleotide correlations that suggest tertiary interactions in the TV-replacement loop-containing mitochondrial tRNAs of the nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Nucleic Acids Res, 1994, 22: 4300–4306
50 Jeon H K, Kim K H, Eom K S. Complete sequence of the mitochondrial genome of *Taenia saginata*: Comparison with *T. solium* and *T. asiatica*. Parasitol Int, 2007, 56: 243–246
51 Jeon H K, Lee K H, Kim K H, et al. Complete sequence and structure of the mitochondrial genome of the human tapeworm, *Taenia asiatica* (Platyhelminthes; Cestoda). Parasitology, 2005, 130: 717–726
52 Jeon H K, Eom K S. *Taenia asiatica* and *Taenia saginata*: Genetic divergence estimated from their mitochondrial genomes. Exp Parasitol, 2006, 113: 58–61
53 Bowles J, McManus D P. Genetic characterization of the *Asian Taenia*, a newly described taeniid cestodes of human. Am J Trop Med Hyg, 1994, 50: 33–44
54 Bowles J, Blair D, McManus D P. Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. Mol Biochem Parasitol, 1992, 54: 165–173
55 Bowles J, Blair D, McManus D P. A molecular phylogeny of the genus *Echinococcus*. Parasitology, 1995, 110: 317–328
56 Bowles J, McManus D P. NADH dehydrogenase 1 gene sequences compared for species and strains of the genus *Echinococcus*. Int J Parasitol, 1993, 23: 969–972
57 Lavikainen A, Lehtinen M J, Meri T, et al. Molecular genetic characterization of the Fennoscandian cervid strain, a new genotypic group (G10) of *Echinococcus granulosus*. Parasitology, 2003, 127: 207–215
58 Le T H, Pearson M S, Blair D, et al. Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*. Parasitology, 2003, 124: 97–112
59 Thompson R C A, McManus D P. Towards a taxonomic revision of the genus *Echinococcus*. Trends Parasitol, 2002, 18: 452–457
60 Thompson R C A. The taxonomy, phylogeny and transmission of *Echinococcus*. Exp Parasitol, 2008, 119: 439–446
61 Romig T. Epidemiology of echinococcosis. Langenbeck's Arch Surg, 2003, 388: 209–217
62 Scott J C, Stafaniak J, Pawlowski Z S, et al. Molecular genetic analysis of human cystic hydatid cases from Poland: Identification of a new genotypic group (G9) of *Echinococcus granulosus*. Parasitology, 1997, 114: 37–43

63 Moks E, Jõgisalu I, Valdmann H, et al. First report of *Echinococcus granulosus* G8 in Eurasia and a reappraisal of the phylogenetic relationships of 'genotypes' G5-G10. Parasitology, 2008, 135: 647–654

64 Nakao M, McManus D P, Schantz P M, et al. A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. Parasitology, 2007, 134: 713–722

65 Hüttner M, Nakao M, Wassermann T, et al. Genetic characterization and phylogenetic position of *Echinococcus felidis* (Cestoda: Taeniidae) from the African lion. Int J Parasitol, 2008, 38: 861–868

66 Blouin M S, Yowell C A, Courthey C H, et al. *Haemonchus placei* and *Haemonchus contortus* are distinct species based on mtDNA evidence. Int J Parasitol, 1997, 27: 1383–1387

67 Park J K, Kim K H, Kang S, et al. A common origin of complex life cycles in parasitic flatworms: Evidence from the complete mitochondrial genome of *Microcotyle sebastis* (Monogenea: Platyhelminthes). BMC Evol Biol, 2007, 7: 11

68 Justin J L. Non-monophyly of the monogeneans? Int J Parasitol, 1998, 28: 1653–1657

69 Nakao M, Xiao N, Okamoto M, et al. Geographic pattern of genetic variation in the fox tapeworm *Echinococcus multilocularis*. Parasitol Int, 2009, 58: 384–389

70 Xiao N, Qiu J M, Nakao M, et al. *Echinococcus shiquicus* n. sp., a taeniid cestode from Tibetan fox and plateau pika in China. Int J Parasitol, 2005, 35: 693–701

71 Nakao M, Li T Y, Han X M, et al. Genetic polymorphisms of *Echinococcus* tapeworms in China as determined by mitochondrial and nuclear DNA sequences. Int J Parasitol, 2010, 40: 379–385

72 Eckert J, Gemmell M A, Meslin F X, et al. WHO/OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern. Paris: World Organisation for Animal Health (Office International des Epizooties) and World Organisation for Animal Health, 2001

73 Hancock K, Broughel D E, Moura I N, et al. Sequence variation in the cytochrome oxidase I, internal transcribed spacer 1, and Ts14 diagnostic antigen sequences of *Taenia solium* isolates from South and Central America, India, and Asia. Int J Parasitol, 2001, 31: 1601–1607

74 Nakao M, Okamoto M, Sako Y, et al. A phylogenetic hypothesis for the distribution of two genotypes of the pig tapeworm *Taenia solium* worldwide. Parasitology, 2002, 124: 657–662

75 Martinez-Hernandez F, Jimenez-Gonzalez D E, Chenillo P, et al. Geographical widespread of two lineages of *Taenia solium* due to human migrations: Can population genetic analysis strengthen this hypothesis? Infect Genet Evol, 2009, 9: 1108–1114

76 Zhan B, Li T, Zhang F, et al. Species-specific identification of human hookworms by PCR of the mitochondrial cytochrome oxidase I gene. J Parasitol, 2001, 87: 1227–1229

77 Ito A, Nakao M, Wandra T, et al. Taeniasis and cysticercosis in Asia and the Pacific: Present state of knowledge and perspectives. Southeast Asian J Trop Med Public Health, 2005, 36: 123–130

78 Eom K S, Jeon H K, Kong Y, et al. Identification of *Taenia asiatica* in China: Molecular, morphological, and epidemiological analysis of a Luzhai isolate. J Parasitol, 2002, 88: 758–764

79 Anantaphruti M T, Yamasaki H, Nakao M, et al. Sympatric occurrence of *Taenia solium*, *T. saginata*, and *T. asiatica*, Thailand. Emerg Infect Dis, 2007, 13: 1413–1416

80 Murell K D. Epidemiology of taeniosis and cysticercosis. In: Murell K D, ed. WHO/FAO/OIE Guidelines for the Surveillance, Prevention and Control of Taeniosis/Cysticercosis. Paris: World Organisation for Animal Health, 2005. 27–44

81 Mayta H, Gilman R H, Prendergast E, et al. Nested PCR for the specific diagnosis of *Taenia solium* taeniasis. J Clin Microbiol, 2008, 46: 286–289

82 Yamasaki H, Allan J C, Sato M O, et al. DNA differential diagnosis of taeniasis and cysticercosis by multiplex PCR. J Clin Microbiol, 2004, 42: 548–553

83 Yamasaki H, Nakao M, Sako Y, et al. Mitochondrial DNA diagnosis for taeniasis and cysticercosis. Parasitol Int, 2006, 55: S81–S85

84 Jeon H K, Chaib J Y, Kong Y, et al. Differential diagnosis of *Taenia asiatica* using multiplex PCR. Exp Parasitol, 2009, 121: 151–156

85 Yamasaki H, Nakao M, Sako Y, et al. DNA differential diagnosis of human taeniid cestodes by base excision sequence scanning thymine-base reader analysis with mitochondrial genes. J Clin Microbiol, 2002, 40: 3818–3821

86 Nkouawa A, Sako Y, Nakao M, et al. Loop-mediated isothermal amplification method for differentiation and rapid detection of *Taenia* species. J Clin Microbiol, 2009, 47: 168–174