A single nucleotide polymorphism map of the mitochondrial genome of the parasitic nematode *Cooperia oncophora*

M. VAN DER VEER *and* E. DE VRIES*

Division of Parasitology and Tropical Veterinary Medicine, Department of Infectious Diseases and Immunology, Utrecht University, PO Box 80.165, 3508 TD Utrecht, The Netherlands

(*Received 19 August 2003; revised 3 October 2003; accepted 4 October 2003*)

SUMMARY

The 13 636 bp mitochondrial (mt) genome sequence of the trichostrongylid nematode *Cooperia oncophora* was determined.† Like the mt genomes of other nematodes it is AT rich (76·75%) and cytidine is the least common nucleoside in the coding strand. There are 2 ribosomal RNA (*rrn*) genes, 22 transfer RNA (*trn*) genes and 12 protein coding genes. The relatively short AT-rich region (304 bp) and the lack of a non-coding region between two of the NADH dehydrogenase genes, *nad*3 and *nad*5, makes the mt genome of *C. oncophora* one of the smallest known to date, having only 525 bp of non-coding regions in total. The majority of the *C. oncophora* protein encoded genes are predicted to end in an abbreviated stop codon like T or TA. In total, 426 single nucleotide polymorphisms (SNP) were mapped on the mt genome of *C. oncophora*, which is an average of 1 polymorphism per 32 bp. The most common SNPs in the mt genome of *C. oncophora* were G/A (59·2%) and C/T (28·4%) transitions. Synonymous substitutions (86·4%) were favoured over non-synonymous substitutions. However, the degree of sequence conservation between individual protein genes of different parasitic nematode species did not always correspond to the relative number of non-synonymous SNPs. The mt genome sequence of *C. oncophora* presents the first mt genome of a member of the Trichostrongyloidea and will be of importance in refining phylogenetic relationships between nematodes. The, still limited, SNP map presented here provides a basis for obtaining insight into the genetic diversity present in the different protein coding genes, *trn*, *rrn* and non-coding regions. A more detailed study of the more variable regions will be of use in determining the population genetic structure of *C. oncophora*. Ultimately this knowledge will add to the understanding of the host–parasite relationship.

Key words: *Cooperia oncophora*, parasitic nematode, mitochondrial genome, single nucleotide polymorphism, intraspecific genetic variation.

INTRODUCTION

Cooperia oncophora, an economically important parasitic nematode of cattle, belongs to the superfamily of Trichostrongyloidea. Studies of population structures at the genetic level of several strongylid species have given insight into host–parasite relations (Blouin *et al*. 1995). Mitochondrial (mt) DNA, in particular, is very informative in studying aspects like population history, population size and population subdivision (Blouin *et al*. 1992; Tarrant *et al*. 1992; Dame, Blouin & Courtney, 1993) owing to its strict maternal inheritance, high mutation rate and absence of recombination (Gyllensten, Wharton & Wilson, 1985; Anderson *et al*. 1995).

The metazoan mt genome is circular and varies in size between 14 and 18 kb (Wolstenholme, 1992; Boore, 1999). For parasitic nematodes the mt genome usually encodes 12 proteins (*atp*6, *cob*, *cox*1-3 and *nad*1-4, 4L, 5-6) which are the components of the respiratory chain enzyme complexes (Okimoto *et al*. 1992; Keddie, Higazi & Unnasch, 1998; Hu, Chilton & Gasser, 2002), with the exception of *Trichinella spiralis* which, in addition, encodes a putative *atp*8 gene (Lavrov & Brown, 2001). Additionally the mt genome codes for 22 transfer RNAs (*trn*) and 2 ribosomal subunit RNAs (*rrn*). It generally contains at least 1 non-coding region which is assumed to have a function in the regulation of transcription and control of DNA replication of the mt genome (Clayton, 1982). Mt genomes have been sequenced from *Caenorhabditis elegans*, *Ascaris suum* (Okimoto *et al*. 1992), *Onchocerca volvulus* (Keddie *et al*. 1998), *Trichinella spiralis* (Lavrov & Brown, 2001), *Ancylostoma duodenale* and *Necator americanus* (Hu *et al*. 2002). The mt genome of *C. oncophora* presents the first member of the Trichostrongyloidea and its study can contribute to a further understanding of nematode evolution.

One of the aspects making mtDNA a good marker for studying population structures is the higher rate of evolution in mt genomes compared to the nuclear genome, probably caused by frequent exposure of mtDNA to reactive oxygen metabolites (Wolstenholme, 1992; Boore, 1999; Wallace, 1999).

* Corresponding author. Tel: +31 30 253 2582. Fax: +31 30 254 0784. E-mail: E.Vries@vet.uu.nl

† Sequence data have been deposited in the EMBL/GenBank Data libraries under Accession number AY265417. The single nucleotide polymorphisms have been deposited in the dbSNP Database (http://www.ncbi.nlm.nih.gov/SNP) under ID numbers ss8485731-ss8486156.

DOI: 10.1017/S0031182003004633 Printed in the United Kingdom

Therefore, the mt genome of *C. oncophora* was chosen as a target for identifying molecular markers since the higher rate of base substitution facilitates the characterization of single nucleotide polymorphisms (SNPs). To identify regions apparently enriched in variable positions an approach was taken whereby each position on the mt genome was determined from at least 3 different individuals. The generated data will provide a better view of the genetic diversity present in *C. oncophora* populations and ultimately contribute to dissecting the genetic population structure.

MATERIALS AND METHODS

Parasites and DNA isolation

Adult parasites were obtained from the intestine of female Holstein Friesian calves 28 days after infection (p.i.) with 100 000 infective *C. oncophora* larvae (L3). Before infection random samples of larvae were microscopically verified as being *C. oncophora*. Contamination with heterologous species of trichostrongylids was ruled out because samples of larvae were differentiated to the generic level by microscopical examination (MAFF, 1986). DNA used as template in the PCR was obtained from a pool of worms and individual *C. oncophora* adults that were also carefully checked on the basis of generally accepted morphological criteria (MAFF, 1986). Total DNA was isolated by proteinase K digestion and phenol extraction as described previously (Roos *et al.* 1990). To confirm that the isolated DNA was from *C. oncophora* a *Cooperia*-specific COX1 PCR was performed using the primers COX1FN 5′TAATGCCTAGTATAAT(C/T)GGTGGTTT′3 and COX1RN 5′CCCAGCTAAAACAGGTAAAGATAAT′3.

Long template PCR amplification

The complete mt DNA was amplified from 5 ng of total genomic DNA isolated from pooled (20 mg) adult worms by means of two overlapping long PCR reactions. COX1FN in combination with NAD6R 5′TTTAAATACAACTTTACTCCTGCTCTT′3 covered a ∼6 kb region. The second combination of COX1RN and NAD6F 5′CATATTTGGTTTTCTTACTTTATTTG′3 yielded a ∼8 kb product. The primers COX1FN and COX1RN are based on the sequence of the partial *cox*1 gene from *C. oncophora* (Accession numbers: AY229868-AY229873). The NAD6F and R primers are based on the *nad*6 gene sequence of *Ascaris suum* and *Caenorhabditis elegans* (Okimoto *et al.* 1992). The Expand Long Template PCR system (Roche, Mannheim, Germany) was used for amplification of the two fragments. Cycling conditions were according to the manufacturer's protocol with modifications to annealing temperature. In short, 2 min initial denaturation at 92 °C followed by 10 cycles of 30 s denaturation at 92 °C, 30 s annealing at 55–50 °C lowering each cycle with 0·5 °C and 12 min extension at 68 °C, followed by 30 cycles of 30 s denaturation at 92 °C, 30 s annealing at 50–45 °C and 12 min at 68 °C with an additional cycle elongation of 20 s for each cycle and a final cycle of 7 min at 68 °C. The amplicons were digested with *Mbo*I and *Taq*I to produce overlapping clones. The restriction fragments were directly cloned into a compatible digested (*Bgl*II or *Nar*I) pUC-PCR vector (de Vries, 1998). A total of 38 clones, ranging in size from 100–2000 bp, were sequenced on both strands with universal M13 sequencing primers. Since the mt sequence was amplified from a large pool of worms (>5000 individuals), it can be assumed that each clone contained a sequence from a different individual. The AT-rich parts of the mt genome were difficult to clone. Therefore, these parts were amplified and sequenced directly from individual worms. Also the parts of the mt genome that were covered by less than 3 clones were amplified and sequenced directly from individual worms, up to a minimum of 3 sequences for each region. In total an additional 35 fragments, ranging in size from 309–846 bp, covering the parts with gaps and insufficient number of clones were obtained by long PCR reactions performed on single worms. Each individual worm was used only once. Primers for additional PCR reactions were designed on the sequences derived from the *C. oncophora* mitochondrial clones. Long PCR was performed according to the manufacturer's instructions (see above). Depending on the predicted length of the sequence the elongation time and annealing temperature were adjusted. The complete mt genome was sequenced at least 3 times on both strands.

Sequence analysis

Clones and amplicons were purified using the GFX™ purification kit (Amersham Biosciences). Samples were sequenced by BaseClear B.V. (Leiden, The Netherlands) using terminator chemistry (Perkin Elmer) and an Applied Biosystems 3100 genetic analyser. The LaserGene 5.03 package (DNAStar Inc., Madison, WI, USA) was used for all sequence analyses. The SeqMan program was used for sequence assembly and identification of the SNPs. All SNPs were sequenced on both strands and manually checked by inspection of the trace files. Alignment with the *C. elegans* mt DNA sequence and BLAST analysis were used for identification of the genes. EditSeq was used to translate the protein coding genes using standard genetic codes with the modifications ATA (Met), AGA and AGG (Ser) and TGA (Trp) specific for the nematode mt genetic code (Okimoto *et al.* 1992; Jukes & Osawa, 1993). Determination of the AT content and codon usage was also done with the EditSeq program. MegAlign was used for alignment of the *C. oncophora* amino acid sequences

Table 1. Sequence pair distances in percentage identity from the amino acid sequences of the 12 protein coding genes and the nucleotide sequences of the 2 *rrn* genes, AT-rich non-coding region and the complete mt genome

(Comparison of *C. oncophora* (Co) with *A. duodenale* (Ad), *A. suum* (As), *C. elegans* (Ce), *N. americanus* (Na), *O. volvulus* (Ov) and *T. spiralis* (Ts). The protein coding genes are listed from most conserved between *C. oncophora* and the other nematodes (except *O. volvulus* and *T. spiralis*) to least conserved. The ratio synonymous/non-synonymous SNPs for the protein coding genes of *C. oncophora* is given.)

Gene or region	Co/Ad	Co/As	Co/Ce	Co/Na	Co/Ov	Co/Ts	Ratio
*cox*1	94·5	89·3	88·8	95·4	52·5	56·4	21·0
*cox*2	87·9	80·6	85·7	88·7	40·1	38·2	5·5
*cox*3	91·8	79·2	85·1	90·6	31·4	26·8	9·0
*atp*6	86·4	73·4	77·9	86·4	21·6	13·6	1·5
cob	83·8	74·5	76·8	80·3	50·6	34·9	3·3
*nad*1	79·3	67·6	73·4	77·6	50·3	39·7	11·0
*nad*5	73·6	62·6	67·9	74·0	38·9	25·5	4·67
*nad*4	72·4	61·9	67·0	71·1	44·3	26·2	29·0
*nad*4L	70·1	59·7	67·5	71·4	37·7	21·3	4·0
*nad*3	66·7	62·2	60·4	64·0	36·0	27·0	4·0
*nad*2	59·0	45·0	51·8	60·1	34·9	17·8	3·5
*nad*6	56·9	54·2	52·1	58·3	24·8	14·5	0·67
*rrn*S	84·9	74·7	78·2	81·6	66·2	11·0	—
*rrn*L	80·0	71·2	76·5	77·3	66·6	32·8	—
AT-rich	53·0	60·9	65·1	58·4	50·4	—	—
Genome	79·8	72·0	75·6	78·4	56·7	—	—

with those of *Ancylostoma duodenale*, *A. suum*, *C. elegans*, *Necator americanus*, *Onchocerca volvulus* and *Trichinella spiralis*. Pairwise alignments were performed by the ClustalW slow/accurate method (Thompson, Higgins & Gibson, 1994), calculation of percentage identity between pairs was based on the shortest sequence, with default parameters. The aligned sequences were imported into ClustalX (Thompson *et al*. 1997) and unrooted phenograms were constructed by the Neighbour-Joining (NJ) algorithm using 1000 bootstrap replicates. Phenograms were drawn using NJplot (Perriere & Gouy, 1996). GeneQuest was used to scan non-coding regions for repeats. The mt sequences from the other nematodes were downloaded from Genbank, Accession numbers; *A. duodenale*: AJ417718 and *N. americanus*: AJ417719 (Hu *et al*. 2002), *A. suum*: X54253 and *C. elegans*: X54252 (Okimoto *et al*. 1992), *O. volvulus*: AF015193 (Keddie *et al*. 1998) and *T. spiralis* AF293969 (Lavrov & Brown, 2001). The partial *nad*4 sequences from *Haemonchus placei* AF070825-AF070786, *Haemonchus contortus* AF070785-AF070746, *Teladorsagia circumcincta* AF070916-AF070877 and *Mazamastrongylus odocoilei* AF070876-AF070837 (Blouin *et al*. 1995).

RESULTS

General characteristics

In total 73 fragments (47 161 bp total length, 3·5-fold coverage) were sequenced on both strands and assembled in a single contig. Similar to other nematode mt genomes, the mt genome of *C. oncophora* is AT rich (76·25%) and has an unequal nucleotide composition in the coding strand (29·8% A, 46·94% T, 15·94% G and 6·43% C). The mt genome contains the genes (Table 1) for 12 proteins (3 subunits of cytochrome c oxidase, *cox*1, *cox*2 and *cox*3; 1 subunit of cytochrome c-ubiquinol oxidoreductase, *cob*; 7 subunits of NADH dehydrogenase, *nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5 and *nad*6; and 1 subunit of the ATP synthase, *atp*6); 22 tRNAs (*trns*) and the small and large subunit ribosomal RNAs, *rrn*S and *rrn*L. As in other nematodes, possibly with the exception of *T. spiralis* (Lavrov & Brown, 2001), an *atp*8 gene is absent in *C. oncophora*. The relative order of the genes in the mt genome of *C. oncophora* (Fig. 1) is identical to that in *A. duodenale*, *C. elegans* and *N. americanus* (Okimoto *et al*. 1992; Hu *et al*. 2002). Whereas in *A. suum* the AT-rich region is located between the *trn*S (UCN) and *trn*N genes and in *O. volvulus* and *T. spiralis* also a number of *trn*s and protein coding genes are located differently (Okimoto *et al*. 1992; Keddie *et al*. 1998; Lavrov & Brown, 2001). After correction for the different location of the genes and AT-rich region of *O. volvulus* and *A. suum* the percentage identity is the highest with *A. duodenale* (79·8%), followed by *N. americanus* (78·4%), *C. elegans* (75·6%), *A. suum* (72·0%) and *O. volvulus* (56·7%) (Table 1).

Protein genes and codon usage

The genes of the mt genome of *C. oncophora* are arranged in an extremely economic fashion with a complete absence of intergenic sequences between many gene pairs (Fig. 1). Only 11 intergenic regions were found, the lowest number of intergenic regions identified in a nematode mt genome to date.

https://doi.org/10.1017/S0031182003004633 Published online by Cambridge University Press

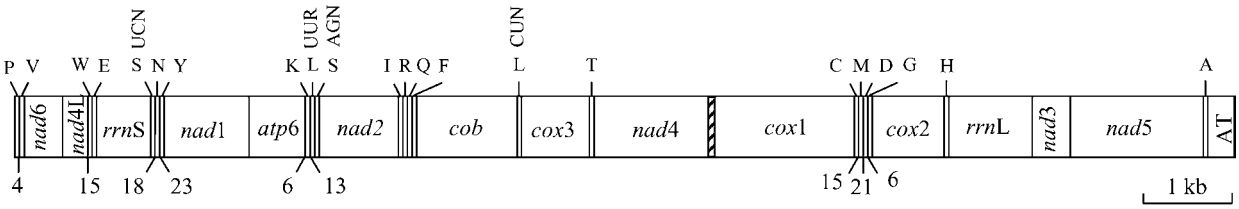

Fig. 1. Schematic representation of the mitochondrial genome of *Cooperia oncophora*. Transcription of the genes is predicted to be all in the same direction (from left to right). AT represents the AT-rich non-coding region. The Long Non-coding (LNC) region is shaded, the length of the other 9 non-coding regions is given below. The *trn* genes are represented by their one letter amino acid code at the top.

Eight out of the 12 protein genes terminate with abbreviated translation stop codons, illustrating the optimal use of sequence (6 times T and 2 times TA, see Table 2). Alternative initiation codons such as ATT, ATA and TTG have been described for other nematode mt genomes (Okimoto *et al*. 1992; Keddie *et al*. 1998; Hu *et al*. 2002). The ATT initiation codon, coding for isoleucine, is used 7 times as a start codon, ATA is used as the start codon for the *cob* gene and the remaining 4 genes, *cox*3, *nad*1, *nad*2 and *nad*4, employ a TTG codon (Table 2).

Table 1 shows the percentage identities for the inferred amino acid sequences of the protein genes from *C. oncophora* compared with *A. duodenale*, *A. suum*, *C. elegans*, *N. americanus*, *O. volvulus* and *T. spiralis*. Except for *cox*1, *T. spiralis* has the lowest similarity scores when compared with *C. oncophora*. The highest identity scores were found for the three cytochrome c oxidase subunits and the lowest identities were found for the *nad*2 and *nad*6 proteins. For *O. volvulus* and *T. spiralis*, this pattern is slightly different with, for instance, relatively well-conserved *nad*1 and *cob* genes and a poorly conserved *atp*6 gene. Fig. 2A presents an unrooted phenogram of the combined amino acid sequences encoded by the mt genome of different nematodes as derived from Neighbour-Joining (NJ) analysis. The tree clearly indicates the relative large phylogenetic distance of *O. volvulus* and *T. spiralis* to the other displayed nematodes. The AT bias and the preference of a G over a C in the coding strand has been established for other nematode mt genomes and is, of course, strongly reflected in the codon usage of *C. oncophora* mt protein coding genes (Table 3). The 10 most used codons (Phe (TTT), Leu (TTA), Ile (ATT), Tyr (TAT), Leu (TTG), Asn (ATT), Met (ATA), Val (GTT), Ser (AGT), Gly (GGT)) are all AT rich, and 7 contained a thymidine in the 3rd position. None of the most frequently used codons contained a cytidine. In contrast, the least used codons (Pro (CCC), Ala (GCC), Arg (CGA and CGG), Asp (GAC), Ile (ATC), Leu (CTA and CTC), Cys (TGC), Gly (GGC), Thr (ACC)) mostly contain a cytidine at the 1st or 3rd position. Only 2 codons were never used, (Ser (TCC) and Arg (CGC)).

tRNA-rRNA genes

Phylogenetic analysis of the *C. oncophora rrn* genes with those of other nematodes resulted in the same order of conservation as observed for the protein-coding genes (Fig. 2B). The overall nucleotide sequences of the *trn*s, which were between 52 and 61 bp, and the predicted secondary structure of the *C. oncophora* tRNAs were similar to those of other nematodes. General features of the *trn*s, like a missing variable loop and TΨC arm, which are replaced by the TV replacement loop, were found for 20 of the *trn*s. The overall secondary structure of 7 bp in the stem structure of the amino-acyl arm is also seen for *C. oncophora*. Mismatches in the amino-acyl arm stem structure were seen for 16 of the 22 *trn*s and most are situated at the base of the stem concerning mostly T/G pairs. All 22 *trn*s have a 5 bp anticodon arm stem combined with a 7 nt loop. As described for other nematodes the anticodon is preceded by a pyrimidine followed by a thymidine and the anticodon is followed by a purine. In only 8 cases a single mismatch in the anticodon stem was found. The consensus structure for the 20 *trn*s having a dihydrouridine (DHU) arm consists of a 4 bp stem structure and a loop varying from 4–10 nt. In 16 out of 20 *trn*s mismatches in the DHU stem were found, overall these are in the first or last bp and all but one were T/G pairs. The 2 *trn*s that differed in basic structure, *trn*S (AGN) and *trn*S (UCN), lack the DHU arm and instead have a variable loop of 4 (UCN) or 6 (AGN) nt. The 2 *trn*Ss also contain a TΨC arm which has a stem structure of 3 bp and a variable loop of 4 nt (AGN) and 5 nt (UCN).

Non-coding regions

The two longest non-coding regions, between *trn*A and *trn*P and between *nad*4 and *cox*1, are believed to be involved in regulation of transcription and control of DNA replication (Clayton, 1991; Okimoto *et al*. 1992). The location of the 304 bp non-coding AT-rich region (85·53% AT) between the *trn*A and *trn*P genes corresponds to its location in *C. elegans*, *N. americanus* and *A. duodenale* (Okimoto *et al*. 1992; Hu *et al*. 2002). The six 43 bp direct repeats

https://doi.org/10.1017/S0031182003004633 Published online by Cambridge University Press

Table 2. Nucleotide position, AT content (%), for the genes and 2 non-coding regions (AT-rich and Long Non-coding (LNC) region between *nad*4 and *cox*1) of the *Cooperia oncophora* mt genome

(The translation initiation and termination codons and the number of amino acids (aa) of the inferred protein coding genes are given. To assist localization of the 9 short non-coding regions, positions preceded by a non-coding region are shown in bold.)

	Nucleotide position				Codon	
Gene or region	First	Last	No. of aa	AT content	Initiation	Termination
*trn*P	1	55		83·64		
*trn*V	**60**	113		77·87		
*nad*6	114	551		79·00	ATT	TA
*nad*4L	552	784	77	80·77	ATT	TAG
*trn*W	**800**	854		89·09		
*trn*E	855	908		79·63		
*rrn*S	909	1604	527	75·57		
*trn*S (UCN)	1605	1657		81·13		
*trn*N	**1676**	1729		75·93		
*trn*Y	**1753**	1807		87·27		
*nad*1	1808	2678	370	72·07	TTG	T
*atp*6	2679	2278	145	77·00	ATT	TAA
*trn*K	**3285**	3345		73·77		
*trn*L (UUR)	**3359**	3413		74·55		
*trn*S (AGN)	3414	3465		69·32		
*nad*2	3466	4300	290	80·82	TTG	T
*trn*I	4301	4360		83·33		
*trn*R	4361	4414		77·78		
*trn*Q	4415	4468		79·63		
*trn*F	4469	4523		80·00		
cob	4524	5634	255	73·77	ATA	T
*trn*L (CUN)	5635	5690		85·71		
*cox*3	5691	6456	232	72·55	TTG	T
*trn*T	6457	6511		80·00		
*nad*4	6512	7741	234	79·84	TTG	TAA
LNC region	7742	7810		78·26		
*cox*1	7811	9390	199	70·79	ATT	TA
*trn*C	9391	9449		79·66		
*trn*M	**9465**	9522		67·24		
*trn*D	**9544**	9598		81·82		
*trn*G	**9605**	9659		74·55		
*cox*2	9660	10358	526	72·53	ATT	TAA
*trn*H	10359	10411		77·36		
*rrn*L	10412	11360		81·88		
*nad*3	11361	11694	278	78·98	ATT	T
*nad*5	11695	13276	77	79·06	ATT	T
*trn*A	13277	13332		78·57		
AT region	13333	13636		85·53		

present in the AT-rich region of *C. elegans* or any other direct repeat could not be detected in *C. oncophora*. Instead, 2 stretches of 8 consecutive AT dinucleotides are present. The second AT stretch has an overlap with an upstream inverted repeat of 18 bp separated by 1 G. Within the second longest non-coding region (69 bp) between the *cox*1 and *nad*4 genes (not present in *O. volvulus*) a hairpin structure could be formed (Fig. 3). The region is shorter as in the 4 other nematodes (Okimoto *et al*. 1992; Keddie *et al*. 1998; Hu *et al*. 2002) and the AT content is lower (78·26%). The other 9 intergenic regions of the *C. oncophora* mt genome vary in size from 4 to 23 bp. Secondary structures or repeated sequences could not be identified within these regions.

Single nucleotide polymorphisms

The mt genome of *C. elegans* appears to have a two orders of magnitude higher substitution rate, $9{\cdot}7 \times 10^{-8}$ ($\pm 2{\cdot}4 \times 10^{-8}$) per site per generation than previously estimated (Denver *et al*. 2000). Furthermore the mt genome has a lack of recombination and is strictly maternally inherited. Consequently sequences derived from the mt genome are an excellent target for identifying molecular markers which can be used for defining the population structure. For this reason variable positions were directly determined by sequencing fragments derived from multiple individuals rather than sequencing the complete mt genome from a single worm. Each position in the mt genome was sequenced from

https://doi.org/10.1017/S0031182003004633 Published online by Cambridge University Press

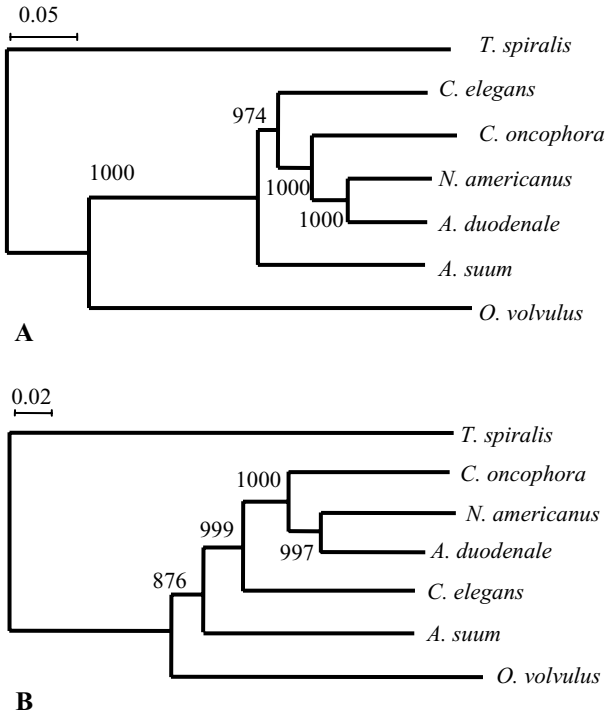

Fig. 2. Neighbour-Joining (NJ) trees derived from mitochondrial data for 7 nematode species. (A) Data derived from the amino acid sequence of the 12 protein coding genes. (B) Data derived from the nucleotide sequences of the *rrn* genes. Bootstrap values are given at the branches.

a minimum of 3 and maximum of 12 different individuals. Table 4 shows the 426 SNPs (1 polymorphism per 32 bp) detected in the 13 636 bp mt genome of *C. oncophora*. No deletions or insertions were identified. A total of 58 SNPs (13·6%) give rise to amino acid substitutions. The *nad*4L gene contains the most SNPs, 1 polymorphism per 23·4 bp and *nad*6 gene the least, 1 per 43·8 bp. The degree of conservation of the individual genes in between the different nematode species (Table 1) is not very well reflected in the distribution of non-synonymous SNPs. For instance, one of the most conserved protein sequences like *atp*6 has 10 non-synonymous SNPs out of a total of 25 whereas the less well-conserved *nad*4 has only 1 out of 30.

A bias for nucleotide transitions (G/A or T/C) over transversions in mtDNA was very distinct in the mt genome of *C. oncophora*. From all SNP positions, 59·2% were harbouring G/A transitions, 28·4% T/C transitions, 7·5% A/T transversions and 4·9% consisted of G/T, C/G and A/C transversions. There is only one case in which 3 different nucleotides were seen (A, C or T) in the same position. Most of the SNPs are synonymous third codon position changes (80·7%). Of the 267 SNPs found in the third codon position only 4 are non-synonymous. First and second codon positions are mainly non-synonymous (54 out of 67). The 12 synonymous first position changes are in Leu or Ser codons.

Table 3. Frequencies (%) of amino acid (aa) codon usage in *Cooperia oncophora* mt genes

(Codons that contained a polymorphic site (2·39%) are not included in the table.)

aa	Codon	Frequency
Phe	TTC	0·12
	TTT	12·70
Leu	TTA	10·43
	TTG	4·70
Leu	CTA	0·06
	CTC	0·03
	CTG	0·20
	CTT	0·12
Ile	ATC	0·06
	ATT	6·75
Met	ATA	4·21
	ATG	1·90
Glu	GAA	1·43
	GAG	0·82
Ser	TCA	1·40
	TCC	0·0
	TCG	0·53
	TCT	2·16
Ser	AGA	2·10
	AGC	0·12
	AGG	0·47
	AGT	3·77
Pro	CCA	0·55
	CCC	0·09
	CCG	0·18
	CCT	1·37
His	CAC	0·15
	CAT	1·34
Gln	CAA	0·79
	CAG	0·38
Tyr	TAC	0·26
	TAT	5·14
Arg	CGA	0·06
	CGC	0·0
	CGG	0·09
	CGT	0·79
Thr	ACA	0·82
	ACC	0·06
	ACG	0·29
	ACT	1·69
Val	GTA	3·13
	GTC	0·15
	GTG	1·26
	GTT	4·03
Asn	AAC	0·12
	AAT	4·47
Lys	AAA	2·39
	AAG	0·79
Asp	GAC	0·06
	GAT	1·72
Ala	GCA	0·91
	GCC	0·06
	GCG	0·23
	GCT	1·72
Cys	TGC	0·03
	TGT	1·17
Gly	GGA	0·73
	GGC	0·03
	GGG	0·73
	GGT	3·56
Trp	TGA	1·72
	TGG	0·35
Stop	TAA	0·12
	TAG	0·03

https://doi.org/10.1017/S0031182003004633 Published online by Cambridge University Press

Fig. 3. Potential hairpin structure found within the long non-coding region between the *cox*1 and *nad*4 genes of *Cooperia oncophora*.

The number of SNPs found within the *trn* sequences is lower than average (1 polymorphism per 60 bp was identified instead of the 1/32 bp for the complete mt genome). This fits with the observation that *trn* sequences are better conserved than the other sequences. In general, within non-coding regions more variation is seen which is also the case in most non-coding regions of the *C. oncophora* mt genome. For instance, in the region between *nad*4L-*trn*W, which is 15 bp in length, 5 SNPs were discovered. The 13 bp region between *trn*K-*trn*L covered 3 SNPs. The one exception is the region between the *nad*4 and *cox*1 genes (69 bp), which is a relatively long non-coding region without a single SNP, suggesting that this region is indeed of importance for the integrity of the mt genome.

DISCUSSION

The 13 636 bp mtDNA sequence of *C. oncophora* is the first reported from a trichostrongylid nematode and fits in size between those of *N. americanus* (13 604 bp) and *A. duodenale* (13 721 bp), the two smallest known mt genomes of nematodes (Hu *et al.* 2002). Although the fast evolution of mtDNA makes it often less useful for determining the phylogenetic relationships between closely related species, aspects like gene-order changes are often highly informative in reconstructing the phylogeny regarding branch-points deeper in the tree. Here, the relative order of genes, the degree of conservation in pairwise comparisons of individual genes and the analysis of combined protein sequences or *rrn*S gene sequence consistently confirm the close relationship between *C. oncophora*, *A. duodenale* and *N. Americanus* (all Strongylida). Moreover, the Strongylida in the NJ tree form a clade with the Rhabditida (*C. elegans*) in the NJ tree as has been shown previously (Blaxter *et al.* 1998). The size reduction of the mt genome of *C. oncophora* in comparison to *C. elegans* and *A. suum* is mainly due to a short AT-rich region (304 bp). The hypervariability of the AT-rich region is clearly reflected by percentage nucleotide identity between the other nematodes and *C. oncophora*. Considering the AT content (85·53%) a percentage identity between 50 and 56% is insignificant. The hairpin structure which can be formed in the second longest non-coding region between the *cox*1 and *nad*4 gene did, in contrast to the hairpin structure in *C. elegans* and *A. suum*, not contain a stretch of Ts which is assumed to be involved in initiating second (L) strand synthesis (Okimoto *et al.* 1992).

The extremely efficient use of mtDNA as reflected in the utilization of abbreviated stop codons has already been described for the other nematodes. Most likely a TAA stop codon is created post-transcriptionally after polyadenylation, as is the case in other organisms (Anderson *et al.* 1981; Ojala, Montoya & Attardi, 1981). For *A. duodenale*, *N. americanus*, *A. suum* and *O. volvulus* abbreviated stop codons were found in genes preceding a *trn* in contrast with *C. elegans* where the genes terminating in T or TA were followed by a protein coding gene (Okimoto *et al.* 1992; Keddie *et al.* 1998; Hu *et al.* 2002). In *C. oncophora* T and TA stop codons are directly followed by either a *trn* or protein coding gene. No distinction could be made between the genes following a T or TA stop codon.

The SNP map of the complete mt genome of *C. oncophora* presented herein is the first of any parasitic nematode. An SNP map will be of use for studies on *C. oncophora* population genetics. The current SNP map covers every nucleotide in a range varying from 3 to 12 individuals. This automatically implies that only SNPs that have been established at a high frequency in the population had a high change to be detected. Many of the singleton polymorphic sites reported here will have been selected by chance and do not necessarily represent polymorphisms with high frequencies. On the other hand, the SNP map does already provide a considerable set of parsimony informative positions as well as non-synonymous substitutions that may be the subject of selection. With the help of this map hypervariable regions may be selected which can be searched more intensively for additional informative positions that can be used in determining population structures. As new technologies to score SNPs in a semi-automated way are being developed, it is the aim to implement a high throughput method for marker-assisted screening of potential selection in populations of *C. oncophora* during infection of the host.

Genetic variation within mitochondrial *nad*4 sequences from different trichostrongylids (*Haemonchus placei*, *Haemonchus contortus*, *Teladorsagia circumcincta*, *Mazamastrongylus odocoilei*) has proven to be extremely useful in determining the genetic population structure. The data derived from the mt sequences illustrated that host mobility has a large effect on the genetic structure of the different nematode species (Blouin *et al.* 1995). Comparing the partial *nad*4 sequences used for delineation of the population genetic structure with the corresponding

https://doi.org/10.1017/S0031182003004633 Published online by Cambridge University Press

Table 4. Single nucleotide polymorphisms (426) identified and mapped in the mitochondrial genome of *Cooperia oncophora*

(For protein coding genes positions are given in bp starting at the first position of their translation initiation codon. For the *trn*s, *rrn*s and non-coding regions the nucleotides were numbered in the same direction as the protein coding genes. The incidence of nucleotides at each variable position is given in parenthesis after the nucleotide position where they occur and reflect the nucleotide sequence provided in the heading of each column (i.e. 75 (1/3) in the C/T column indicates that a position 75 one individual harboured a C while 3 other individuals harboured a T at the specific position). The bold position numbers indicate that the substitution is non-synonymous. The initiation (ini) and termination (ter) coordinates refer to the nucleotide positions in the complete mt genome sequence. The frequency given in parenthesis after the gene or region indicates the average occurrence of SNPs for that specific part of the mt genome.)

Gene or region ini … ter	Substitution						
	C/T	A/G	A/T	A/C/T	G/T	A/C	G/C
*atp*6 (1/24) 2679 … 3278	75 (1/3), **107** (1/3), 108 (1/3), 117 (2/4), 192 (1/5), 525 (1/7)	**73** (1/3), **106** (1/3), **118** (1/5), **139** (5/1), **178** (3/3), 240 (3/3), **265** (1/5), 279 (5/1), 282 (1/5), 303 (2/6), 351 (1/7), 372 (1/7), 429 (7/1), 480 (7/1), **514** (7/1), 570 (7/1)			129 (4/2), **370** (1/7)		**227** (4/2)
*cox*1 (1/35·9) 7811 … 9390	30 (1/3), 63 (1/3), 210 (3/1), 231 (1/3), 396 (4/4), 417 (1/7), 468 (1/7), 522 (1/6), 684 (2/2), 777 (1/4), 849 (2/3), 1392 (1/3), 1491 (1/3), 1533 (1/3)	66 (3/1), 156 (3/1), 171 (3/1), 327 (7/1), 339 (6/2), 405 (1/7), 432 (7/1), 483 (6/2), 489 (7/1), 579 (6/1), 588 (1/5), 852 (4/1), 888 (3/2), 912 (1/4), 978 (4/1), 1014 (3/2), 1020 (4/1), 1098 (3/1), 1125 (3/1), 1335 (2/1), 1446 (1/3), 1467 (1/3), 1473 (1/3), 1476 (1/3), **1576** (2/2)	**497** (1/7), 546 (1/6), 747 (3/2), 1227 (1/2), 1536 (1/3)				
*cox*2 (1/26·9) 9660 … 10358	**149** (1/4), 348 (1/4), 558 (4/1), 633 (2/1)	168 (2/3), 273 (7/1), 276 (6/2), 306 (5/1), 309 (1/5), 318 (2/4), 417 (1/4), 423 (4/1), 441 (3/2), 462 (2/3), 486 (3/2), 519 (4/1), **557** (1/4), 573 (1/4), 600 (1/3), 648 (2/1), **695** (2/1)	141 (1/4), 540 (1/4)		**499** (4/1), 612 (3/1)	270 (5/2)	
*cox*3 (1/25·5) 5691 … 6456	342 (1/5), 349 (1/5), 369 (1/5), 417 (1/5), 435 (5/1), 540 (1/6), 552 (1/5), 612 (2/3), 717 (3/2)	81 (6/1), 138 (3/4), 177 (5/2), 183 (2/5), 195 (5/2), 276 (5/2), 306 (2/5), 339 (3/3), **386** (5/1), 444 (5/1), 492 (4/3), 513 (1/6), 522 (1/6), 543 (6/1), **568** (2/4), 648 (4/1), 759 (2/1)	192 (2/5), 534 (1/6)	33 (4/1/2)	**216** (5/2)		
cob (1/25·8) 4524 … 5634	100 (1/4), **170** (4/1), **220** (1/5), **278** (1/3), 501 (5/2), 559 (1/6), 576 (1/6), 759 (1/6), 768 (1/4), 786 (3/2), 868 (4/1), **926** (1/3), 963 (1/2), 991 (1/4), 1020 (8/1), 1074 (5/2)	**14** (1/3), 21 (3/1), 60 (4/1), 69 (1/4), 72 (2/3), 81 (4/1), 186 (4/1), 201 (1/4), 249 (1/3), **292** (1/3), 342 (2/5), 360 (6/1), 435 (6/1), 552 (6/1), 561 (2/5), 591 (5/2), 795 (4/1), 906 (4/1), **907** (4/1), 933 (1/2), 1017 (4/5)	**268** (1/3), **269** (1/3), 465 (6/1), 1032 (5/4)				306 (1/3), **925** (3/1)
*nad*1 (1/24·2) 1808 … 2678	16 (3/1), 237 (1/3), 249 (1/3), 285 (2/2), 322 (3/1), 474 (1/3), 528 (1/3), 666 (3/1), 699 (3/1), 744 (1/3)	21 (1/3), 33 (1/3), 69 (1/3), 90 (2/2), 117 (3/1), 126 (3/1), 258 (1/3), 291 (3/1), 294 (3/1), 303 (1/3), 357 (3/1), 372 (2/2), 405 (1/3), 429 (3/1), 459 (3/1), **469** (3/1), 546 (1/3), 591 (3/1), 630 (3/1), 636 (1/3), **637** (1/3), 690 (2/2), 705 (3/1), 729 (3/1)	165 (1/3)		**164** (1/3)		

*nad*2 (1/30·9) 3466 … 4300	18 (1/3), 129 (3/1), 174 (1/3), 267 (2/2), 549 (3/1), 564 (1/3), 619 (2/2), 675 (1/3), 807 (1/3)	45 (3/1), 54 (1/3), **55** (1/3), 78 (1/3), 90 (1/3), 111 (3/1), **167** (3/1), 198 (3/1), 420 (2/1), 483 (3/1), **544** (1/3), 672 (3/1), 702 (3/1), **752** (1/3), **763** (3/1), 765 (3/1), 816 (3/1)	**757** (3/1)			
*nad*3 (1/33·4) 11361 … 11694	30 (1/2), **59** (1/2)	21 (1/2), 72 (2/1), 99 (2/1), 174 (1/2), 183 (2/1), 234 (1/2), 273 (2/1)	**81** (2/1)			
*nad*4 (1/41·0) 6512 … 7741	252 (2/1), 453 (1/3), 561 (1/2), 564 (1/2), 604 (1/2), 708 (1/2), 756 (2/1), 840 (1/2), 990 (1/3), 993 (1/3)	21 (2/1), 78 (2/1), 111 (2/1), 147 (1/2), 231 (2/1), 273 (1/2), 441 (1/2), 474 (1/2), 504 (2/1), 585 (2/1), 792 (1/2), 795 (1/2), 831 (2/1), 942 (4/1), 951 (4/1), 996 (3/1), 1059 (3/1), 1143 (2/2)	984 (1/3), **1040** (3/1)			
*nad*4L (1/23·4) 551 … 784	69 (7/1), 173 (2/6), 210 (2/6), 222 (2/6)	126 (7/1), 156 (1/7), 159 (3/5), **192** (7/1), 201 (3/5), **230** (6/2)				
*nad*5 (1/31·0) 11695 … 13276	297 (3/1), 369 (1/3), 435 (1/3), 489 (1/3), 534 (3/1), 585 (3/1), 588 (1/3), 658 (1/3), 729 (1/4), 768 (3/1), 769 (1/3), 849 (3/1), **1061** (5/2), 1098 (1/6), 1111 (1/6), 1119 (5/2), 1155 (1/6), 1207 (1/5), 1236 (5/1), 1332 (3/3), **1339** (1/5)	276 (3/1), 306 (1/3), 315 (3/1), 324 (3/1), 372 (2/2), 492 (3/1), 528 (3/1), 561 (1/3), 609 (2/2), 681 (3/1), 786 (1/3), **911** (2/2), 930 (1/3), **937** (3/1), 975 (1/3), 984 (3/1), 1026 (5/2), 1029 (1/6), 1083 (4/3), 1170 (1/6), **1171** (5/2), **1195** (1/5), 1197 (5/1), **1234** (3/3), 1260 (1/5), **1316** (5/1), 1458 (5/1), 1557 (5/1)	**1497** (4/2)		1416 (5/1)	
*nad*6 (1/43·8) 114 … 550	174 (1/2), **300** (1/4)	93 (1/2), **118** (2/1), 234 (2/2), 255 (3/2)		**17** (2/1), **92** (2/1), **161** (2/1)		**111** (1/2)
*rrn*S (1/46·4) 909 … 1604	394 (7/1)	5 (5/2), 95 (7/1), 194 (6/7), 200 (3/9), 201 (11/1), 237 (11/1), 239 (2/10), 284 (10/1), 390 (7/1), 403 (1/7), 445 (1/6), 545 (6/1)	18 (6/2), 243 (10/2)			
*rrn*L (1/43·1) 10412 … 11360	26 (1/2), 159 (1/7), 269 (6/2), 499 (10/1)	13 (2/1), 84 (2/1), 164 (7/1), 228 (5/3), 257 (4/4), 276 (7/1), 297 (7/1), 316 (7/1), 339 (7/1), 341 (6/2), 574 (1/10), 619 (10/1), 773 (2/1), 856 (1/2), 890 (1/2)	194 (7/1), 201 (7/1), 818 (1/2)			
*trn*H 10359 … 10511			17 (2/2)			
*trn*V 60 … 113			54 (2/1)		32 (1/2)	
*trn*W 800 … 854		24 (7/1)				
*trn*E 855 … 908	45 (1/7)	16 (2/6), 46 (7/1)				
*trn*N 1676 … 1729		16 (1/6)				

Table 4. (*Cont.*)

Gene or region ini … ter	Substitution						
	C/T	A/G	A/T	A/C/T	G/T	A/C	G/C
*trn*K 3285 … 3345		16 (1/7)				15 (7/1)	
*trn*L UUR 3359 … 3413	54 (3/1)	43 (3/1), 46 (1/3)					
*trn*R 4361 … 4414		47 (2/2)					
*trn*Q 4415 … 4468		15 (1/3)					
*trn*F 4469 … 4523		18 (2/2), 49 (3/1)	48 (2/2)				
*trn*M 9465 … 9522		39 (3/1)					
*trn*G 9605 … 9659		48 (4/1)	17 (2/3)				
AT rich 13333 … 13636	98 (2/4), 109 (1/5), 291 (2/2)	46 (2/4), 172 (3/1)	65 (2/4)		132 (3/1), 253 (2/1)	124 (3/1)	
*nad*4L-*trn*W 785 … 799	8 (1/7)	2 (7/1), 5 (4/4), 6 (1/7)	15 (1/7)				
*trn*S-*trn*N 1658…1675			16 (6/1)				
*atp*6-*trn*K 3279 … 3284		1 (3/5)					
*trn*K-*trn*L 3346 … 3358	8 (3/1)	3 (1/3), 7 (3/1)					
*trn*M-*trn*D 9523 … 9543	18 (2/2)	20 (2/2)	2 (2/2)				
*trn*N-*trn*Y 1728 … 1752	12 (1/3)	10 (1/3), 19 (3/1), 23 (3/1)					

sequence in *C. oncophora* revealed that of the 13 SNPs within *C. oncophora* 9 were found at the same position in *T. circumcincta*, 6 in *M. odocoilei*, 5 in *H. contortus*, and 4 in *H. placei*. Regarding this high similarity of several genes from *C. oncophora* with the mt genes of other nematodes and the occurrence of SNPs within these nematodes at the same position as compared to *C. oncophora*, the presented SNP map will be helpful in determining SNPs in the mt genomes of related nematodes.

Harm Ploeger is gratefully acknowledged for supporting this study and critically reading the manuscript. The authors wish to thank Professor Dr A. W. C. A. Cornelissen for his helpful suggestions. This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of Ministry of Economic Affairs (UDG 4889).

REFERENCES

ANDERSON, S., BANKIER, A. T., BARRELL, B. G., DE BRUIJN, M. H., COULSON, A. R., DROUIN, J., EPERON, I. C., NIERLICH, D. P., ROE, B. A., SANGER, F., SCHREIER, P. H., SMITH, A. J., STADEN, R. & YOUNG, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature, London* **290**, 457–465.

ANDERSON, T. J., KOMUNIECKI, R., KOMUNIECKI, P. R. & JAENIKE, J. (1995). Are mitochondria inherited paternally in *Ascaris*? *International Journal for Parasitology* **25**, 1001–1004.

BLAXTER, M. L., DE LEY, P., GAREY, J. R., LIU, L. X., SCHELDEMAN, P., VIERSTRAETE, A., VANFLETEREN, J. R., MACKEY, L. Y., DORRIS, M., FRISSE, L. M., VIDA, J. T. & THOMAS, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature, London* **392**, 71–75.

BLOUIN, M. S., DAME, J. B., TARRANT, C. A. & COURTNEY, C. H. (1992). Unusual population genetics of a parasitic nematode; mtDNA variation within and among populations. *Evolution* **46**, 470–476.

BLOUIN, M. S., YOWELL, C. A., COURTNEY, C. H. & DAME, J. B. (1995). Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* **141**, 1007–1014.

BOORE, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research* **27**, 1767–1780.

CLAYTON, D. A. (1982). Replication of animal mitochondrial DNA. *Cell* **28**, 693–705.

CLAYTON, D. A. (1991). Replication and transcription of vertebrate mitochondrial DNA. *Annual Review of Cell Biology* **7**, 453–478.

DAME, J. B., BLOUIN, M. S. & COURTNEY, C. H. (1993). Genetic structure of populations of *Ostertagia ostertagi*. *Veterinary Parasitology* **46**, 55–62.

DE VRIES, E. (1998). pUCPCR1. A vector for direct cloning of PCR products in a double Xcm1 restriction site offering compatible single 3′-overhanging T residues. *Molecular Biotechnology* **10**, 273–274.

DENVER, D. R., MORRIS, K., LYNCH, M., VASSILIEVA, L. L. & THOMAS, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344.

GYLLENSTEN, U., WHARTON, D. & WILSON, A. C. (1985). Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. *Journal of Heredity* **76**, 321–324.

HU, M., CHILTON, N. B. & GASSER, R. B. (2002). The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). *International Journal for Parasitology* **32**, 145–158.

JUKES, T. H. & OSAWA, S. (1993). Evolutionary changes in the genetic code. *Comparative Biochemistry and Physiology, B* **106**, 489–494.

KEDDIE, E. M., HIGAZI, T. & UNNASCH, T. R. (1998). The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Molecular and Biochemical Parasitology* **95**, 111–127.

LAVROV, D. V. & BROWN, W. M. (2001). *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAS and has a gene arrangement relatable to those of coelomate metazoans. *Genetics* **157**, 621–637.

MAFF (1986). *Manual of Veterinary Parasitological Laboratory Techniques*. (Technical Bulletin No. 18, HMSO, London.)

OJALA, D., MONTOYA, J. & ATTARDI, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature, London* **290**, 470–474.

OKIMOTO, R., MACFARLANE, J. L., CLARY, D. O. & WOLSTENHOLME, D. R. (1992). The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* **130**, 471–498.

PERRIERE, G. & GOUY, M. (1996). WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**, 364–369.

ROOS, M. H., BOERSEMA, J. H., BORGSTEEDE, F. H. M., CORNELISSEN, J., TAYLOR, M. & RUITENBERG, E. J. (1990). Molecular analysis of selection for benzimidazole resistance in the sheep parasite *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **43**, 77–88.

TARRANT, C. A., BLOUIN, M. S., YOWELL, C. A. & DAME, J. B. (1992). Suitability of mitochondrial DNA for assaying interindividual genetic variation in small helminths. *Journal of Parasitology* **78**, 374–378.

THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

WALLACE, D. C. (1999). Mitochondrial diseases in man and mouse. *Science* **283**, 1482–1488.

WOLSTENHOLME, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *International Review of Cytology* **141**, 173–216.