BRILL

Nematology 0 (2021) 1-18

Nematology
brill.com/nemy

Characterisation of the complete mitochondrial genome of the insect-parasitic nematode *Heterorhabditis bacteriophora*: an idiosyncratic gene order and the presence of multiple long non-coding regions

Sassia Omar REGEAI [1,*], David A. FITZPATRICK [2,**], Ann M. BURNELL [2] and Thomais KAKOULI-DUARTE [3,***]

[1] *Department of Life Sciences, School of Basic Science, Libyan Academy of Postgraduate Studies, Janzour, Libya*
[2] *Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland*
[3] *Department of Science and Health, enviroCORE, Institute of Technology Carlow, Carlow, Ireland*

Received: 13 October 2020; revised: 24 April 2021
Accepted for publication: 28 April 2021

Summary – We present here the complete mtDNA genome (mitogenome) of *Heterorhabditis bacteriophora*, an important biological control agent of soil-dwelling insect pests in agriculture and horticulture. This is the first description of a mitogenome for a member of the family Heterorhabditidae. The genome contains the typical chromadorean complement of 12 protein-coding genes, 22 tRNA genes and two rRNA genes. All genes are transcribed in the same direction and have a nucleotide composition high in A and T. For the entire genome, the nucleotide contents are 47.02% (T), 28.81% (A), 16.10% (G), 8.08% (C) and 75.83% (AT). *Heterorhabditis bacteriophora* has a unique, idiosyncratic gene arrangement. It differs from that of *Caenorhabditis elegans* in having a block of seven genes: *trnQ-trnF-cytb-trnL1-cox3-trnT-nad4* translocated to a position between *nad3* and *nad5*, as well as having a change in the position of the four tRNA block gene cluster, *trnC-trnM-trnD-trnG*, where *trnC* and *trnM* have switched places and *trnD* and *trnG* have translocated between *nad4* and *nad5* genes. The *H. bacteriophora* mitogenome is 18 128 bp long, and thus is *ca* 4 kb larger than the mitogenomes of most chromadoreans. This relatively large genome is due to the presence of five non-coding regions (NCR): NCR1 (114 bp), NCR2 (159 bp), NCR3 (498 bp), NCR4 (1917 bp) and NCR5 (2154 bp), which make up 26.7% of the genome. The NCR5 had the highest A + T content of 83.47% indicating that this region is the likely AT-rich control region. The complete 498 bp NCR3 sequence is duplicated in NCR4 and in NCR5 (the putative AT-rich control region). Such an organisation has not been reported previously in nematode mtDNA.

Keywords – concerted evolution, entomopathogenic nematode, gene arrangement, genome size, mitogenome, mtDNA.

Soil-dwelling entomopathogenic nematodes (EPN) of the families Heterorhabditidae and Steinernematidae are widely studied because of their importance as biological control agents for insect pests (Poinar, 1979; Burnell & Stock, 2000; Lacy & Georgis, 2012; Labaude & Griffin, 2018). The free-living infective juveniles (IJ) of these nematodes actively seek out insect hosts, penetrate into the insect haemocoel and release cells of a symbiotic bacterium. Susceptible hosts are killed rapidly by septicaemia and the IJ mature to adulthood. Enzymes secreted by the symbiotic bacteria digest the host cadaver into a semi-liquid food source for the nematodes. Nematode growth and reproduction is prolific, giving rise to successive generations until the food resources of the cadaver are exhausted. Then new IJ are formed that disperse from the cadaver into the soil seeking new hosts. Within approximately 2 weeks of infection, up to half a million IJ (g insect)$^{-1}$ are produced (Akhurst & Bedding, 1986).

The genus *Heterorhabditis* and family Heterorhabditidae were erected by Poinar (1976), with *Heterorhabditis bacteriophora* as the type species. The Heterorhabdtitidae have a worldwide distribution, and currently there

* ORCID: https://orcid.org/0000-0001-5077-9014
** ORCID: https://orcid.org/0000-0001-7345-6998
*** Corresponding author, e-mail: thomae.kakouli@itcarlow.ie

DOI 10.1163/15685411-bja10110

are 16 confirmed *Heterorhabditis* species (Bhat *et al*., 2020). *Heterorhabditis bacteriophora* is closely related to the model nematode *Caenorhabditis elegans* (Blaxter *et al*., 1998). It can be cryopreserved (Nugent *et al*., 1996) and cultured *in vitro* on agar plates (Dunphy & Webster, 1989), as well as in large scale liquid culture for commercial application (Ehlers, 2001). The genome of *Photorhabdus luminescens*, the bacterial symbiont *H. bacteriophora*, has been sequenced (Duchaud *et al*., 2003), as has the complete nuclear genome of *H. bacteriophora* (Bai *et al*., 2013) and an improved annotation of the *H. bacteriophora* genome was published by McLean *et al*. (2018). However, information on the mitochondrial genomes within the genus *Heterorhabditis* is lacking (Blouin, 1998, 2002; Liu *et al*., 1999). To date, no description of the complete mitochondrial genome of any species of *Heterorhabditis* has been published.

The mitochondrial genome (mitogenome) of animals is a circular, double-stranded DNA molecule, typically 16-20 kb in length, comprising 36 or 37 genes encoding 22 tRNAs, two ribosomal RNAs (small and large subunits rRNAs (rrnS and rrnL)), 12 or 13 proteins, and a non-coding control region that initiates replication and transcription (Wolstenholme, 1992; Boore, 1999). The encoded proteins are involved in the synthesis of ATP through oxidative phosphorylation, and thus they play a critical role in the metabolic activity of eukaryotic cells. There are no introns (except in Cnidaria), and intergenic spacers are absent or very small (Boore, 1999). The mitogenome has been the most widely studied region of animal genomes, addressing questions of animal genetic diversity (Blouin, 1998), population structure, dynamics and phylogeography (Moritz *et al*., 1987; Harrison, 1989; Derycke *et al*., 2013), evolution and phylogenetics (Avise, 1986; Liu *et al*., 2013; Sultana *et al*., 2013; Sun *et al*., 2014; DeSalle *et al*., 2017). In addition, genome size, gene arrangement, gene number and structure can be easily investigated for comparative mitochondrial genomic and evolutionary studies (Boore & Browne, 1998; Gissi *et al*., 2008; Liu *et al*., 2013; Sultana *et al*., 2013).

We present here the first description of the sequence and structure of the mitochondrial genome of *H. bacteriophora*. This genome is atypically large, due to the presence of multiple non-coding regions that make up 26.7% of the genome, and it has an idiosyncratic gene order. The sequence information presented here will be an important resource in designing primers to study the population genetic structure and for the molecular identification of *Heterorhabditis* species. The unusual features of the *H. bacteriophora* genome also provide an important contribution to studies of the molecular and evolutionary biology of nematode mitochondrial genomes.

Materials and methods

DNA EXTRACTION

Heterorhabditis bacteriophora strain HP88 (Poinar & Georgis, 1990) was used to infect the last instar larvae of the wax moth, *Galleria mellonella*, and the emerging IJ were harvested using White (1927) traps. Pools of IJ were concentrated by centrifugation and approximately 1 ml of IJ pellet was ground in liquid nitrogen and total genomic DNA was prepared by the conventional phenol/chloroform extraction and ethanol precipitation method (Sambrook *et al*., 1989). This DNA was used as a substrate in a long polymerase chain reaction (Long-PCR) to amplify mtDNA fragments.

PCR AMPLIFICATION, CLONING AND SEQUENCING

The entire mitochondrial genome of *H. bacteriophora* was amplified in six overlapping fragments (Supplementary Fig. S1) by long-PCR (Expand 20 Kb Kit, Roche) from total genomic DNA using the oligonucleotide primer sets presented in Supplementary Table S1. These primers were designed using the following sequence information: EST data (Dolan, 2001), partial sequences for the *rrnL*, *cox2* (Joyce *et al*., 1994) and *nad4* (Liu *et al*., 1999) genes, and mtDNA sequences obtained from PCR fragments amplified as part of this study. Long PCR was performed according to the manufacturer's instructions. The elongation times and annealing temperatures were adjusted in accordance with the predicted lengths of the target sequences (Supplementary Table S1). Amplicons were visualised using 0.8% agarose gels. Selected amplicons were purified using DNA Minispin Columns (Millipore, DNA Gel Extraction Kit) and cloned into TOPO pCR2.1 (fragments < 3 kb) or TOPO XL (fragments > 3 kb) vectors (Invitrogen), which were then transformed into TOPO 10 competent *Escherichia coli* cells (Invitrogen) according to the manufacturer's instructions. The cloned fragments were sequenced at Agowa (https://www.nucleics.com/DNA_sequencing_support/sequencing-service/agowa.html) using the primer walking strategy. Both strands were completely sequenced from two independent clones for each mitochondrial fragment to generate an accurate consensus sequence (contig).

The final consensus sequence was assembled manually and with the aid of the CAP3 sequence assembly program (http://doua.prabi.fr/software/cap3).

Sequence analysis, assembly and annotation

Sequences were aligned using ClustalX (Thompson *et al*., 1997). Protein encoding genes were identified by sequence similarity of the translated open reading frames to *C. elegans* mtDNA sequences available in WormBase (https://wormbase.org/tools/blast_blat) using BLASTx searches. The online program Transeq (http://www.ebi.ac.uk/emboss/transeq/) was used to translate protein coding genes using the invertebrate mitochondrial genetic code. The codon usage data for protein coding genes were obtained using the Countcodon program version 4 (http://www.kazusa.or.jb/codon/countcodon.html). The two rRNA genes were identified based on their similarity to *C. elegans* and other nematode mitochondrial rRNA genes at GenBank NCBI (http://www.ncbi.nlm.nih.gov/entrez) and the European Ribosomal RNA database (http://bioinformatics.psb.ugent.be/webtools/rRNA/). The tRNAs were identified using the tRNAscan-SE v.1.21 (Lowe & Eddy, 1997) available at (http://lowelab.ucsc.edu/tRNAscan-SE/). Codons encoding arginine tRNA and the two serine tRNAs, not detected by tRNAscan, were identified using the Dual Organeller GenoMe Annotator (DOGMA) program (Wyman *et al*., 2004) at https://dogma.ccbb.utexas.edu.

The sequences of the AT-rich region and non-coding regions were scanned for the presence of: motifs using the online program 'Yet Another Digging for DNA Motifs Gibbs Sampler' (SeSiMCMC) (Favorov *et al*., 2005) available at http://favorov.bioinfolab.net/SeSiMCMC/ and the MEME-motif discovery tool, version 3.50 (Bailey & Elkan, 1994), available at http://meme-suite.org/doc/meme.html?man_type=web. The 'Tandem Repeats Finder' software (Benson, 1999) available at https://tandem.bu.edu/trf/trf.html and the 'Approximate Tandem Repeat Hunter' program (Wexler *et al*., 2005) available at http://www.cs.technion.ac.il/labs/cbl/atrhunter/ATRinformation.htm were used to detect tandem repeat sequences. Prediction of potential secondary structures was performed using the online Mfold web server at http://unafold.rna.albany.edu/?q=mfold/DNA-Folding-Form (Zuker, 2003). Mitochondrial genomes were visualised using OGDRAW (v1.3.1) (Greiner *et al*., 2019).

Results and discussion

Genome organisation and nucleotide composition

The complete *H. bacteriophora* mitogenome (accession number NC_008534) contains 18 128 bp, which is 4334 bp larger than that of *C. elegans* and it is larger than most other chromadorean mitogenomes, which generally range in size from 13 to 15 kb (Supplementary Table S2). Five non-coding regions (NCR) account for 26.7% of the genome and contribute to its relatively large size.

The genome contains 36 of the 37 genes typical of metazoans (Fig. 1) and these genes are all transcribed in the same direction, as is characteristic of nematodes. There are 12 genes encoding the following proteins: the NADH dehydrogenase subunits *nad1-6* and *nad4L*, cytochrome oxidase subunits *cox1-3*, cytochrome b apoenzyme (*cytb*), and ATP synthase subunit 6 (*atp6*). The remaining genes comprise two rRNA genes (*rrnL* and *rrnS*), together with 22 tRNA genes. These genes are similar in length to their counterparts in other nematodes and are separated by few (2-18 bp) or no bases (Table 1). The *atp8* gene is absent, as is typical of nematodes generally, with the exception of enoplids such as *Trichinella spiralis* (Lavrov & Brown, 2001), *Trichuris discolor* (Liu *et al*., 2012) and *Trichinella nelsoni* (Mohandas *et al*., 2014). The *H. bacteriophora* mitogenome is highly AT-rich, a pattern that is characteristic of nematode and other invertebrate mtDNA genomes (Albu *et al*., 2008). Of the non-coding regions, NCR5 had the highest AT content (83.47%), indicating that this region is the likely AT-rich control region.

Codon usage and amino acid composition

Codon usage in the 12 protein-encoding genes is shown in Table 2. The most common codons are TTT (Phe 13.7%), TTA (Leu 10.07%), ATT (Ile 6.57%), GTT (Val 4.79%), and TAT (Tyr 4.76%). The least common codons are CGG (Arg 0.03%), CAC (His 0.06%), ACC (Thr 0.09%), GCC (Ala 0.12%), and CCC (Pro 0.15%). The greatly reduced frequencies of G and C at the third codon position appear to reflect the A + T mutational bias in nematode mtDNA (Blouin *et al*.,1998) and the reduced selection pressure on third codon positions (Sharp & Matassi, 1994). Two codons were not observed in this mtDNA genome: TGC (Cys) and CGC (Arg).

The inferred amino acid sequences for 12 mtDNA protein coding genes and the nucleotide sequences of the

Table 1. Summary of the *Heterorhabditis bacteriophora* mt genome showing the position of mitochondrial genes, their nucleotide (nt) lengths, their amino acid (aa) lengths, translation initiation and termination codons as well as non-coding regions (NCR).

Gene	Mt position		Sequence length		Codon	
	From	To	nt	aa	Start	Stop
trnP	1	54	54			
NCR1	55	168	114			
trnV	169	222	54			
nad6	223	657	435	144	TTG	TAA
intergenic	658	659	2			
nad4L	660	896	237	78	TTG	TAG
trnW	897	953	57			
intergenic	954	957	4			
trnE	958	1013	56			
rrnS	1014	1714	701			
trnS UCN	1715	1771	57			
trnN	1772	1825	54			
intergenic	1826	1827	2			
trnY	1828	1883	56			
nad1	1884	2755	872	289	TTG	TA
atp6	2756	3352	597	198	ATT	TAA
intergenic	3353	3370	18			
trnK	3371	3433	63			
intergenic	3434	3439	6			
trnL UUR	3440	3494	55			
trnS AGN	3495	3548	54			
nad2	3549	4396	848	282	TTG	TA
trnI	4397	4456	60			
trnR	4457	4512	56			
NCR2	4513	4671	159			
cox1	4672	6258	1587	528	ATT	TAG
intergenic	6259	6263	5			
trnM	6264	6324	61			
intergenic	6325	6337	13			
trnC	6338	6392	55			
intergenic	6392	6393	1			
cox2	6394	7089	696	231	ATG	TAG
intergenic	7090	7095	6			
trnH	7096	7149	54			
rrnL	7150	8114	965			
nad3	8115	8453	339	112	TTG	TAG
NCR3a	8454	8951	498			
trnQ	8952	9006	55			
intergenic	9007	9010	4			
trnF	9011	9066	56			
cytb	9067	10 179	1113	370	TTG	TAG
intergenic	10 180	10 183	4			
trnL CUN	10 184	10 240	57			
cox3	10 241	11 008	768	255	ATT	TAA
intergenic	11 008	11 009	1			
trnT	11 010	11 065	56			
nad4	11 066	12 294	1229	409	ATA	TA
trnD	12 295	12 349	55			

Table 1. (Continued.)

Gene	Mt position		Sequence length		Codon	
	From	To	nt	aa	Start	Stop
trnG	12 350	12 404	56			
NCR4	12 405	14 321	1917*			
NCR3b	13 848	14 345	498			
nad5	14 322	15 905	1584	527	ATT	TAA
intergenic	15 906	15 918	13			
trnA	15 919	15 974	56			
NCR5 (AT-rich region)	15 975	18 128	2154			
NCR3c	16 072	16 569	498			

* Approx. 24 bp overlap with *nad5*.

Table 2. Codon usage in the 12 protein encoding genes of the *Heterorhabditis bacteriophora* mitochondrial genome.

Amino acid	Codon	No.	%	Amino acid	Codon	No.	%
Phe (F)	TTT	451	13.17	Asn (N)	AAT	136	3.97
	TTC	24	0.70		AAC	7	0.2
Leu (L2)(UUR)	TTA	345	10.07	Lys (K)	AAA	57	1.66
	TTG	153	4.47		AAG	44	1.29
Leu (L1) (CUN)	CTT	29	0.85	Tyr (Y)	TAT	163	4.76
	CTC	1	0.03		TAC	8	0.23
	CTA	12	0.35		TAA*		
	CTG	2	0.06		TAG*		
Ile (I)	ATT	225	6.57	Asp (D)	GAT	70	2.04
	ATC	9	0.26		GAC	4	0.12
Met (M)	ATA	129	3.77	Glu (E)	GAA	41	1.20
	ATG	60	1.75		GAG	35	1.02
Val (V)	GTT	164	4.79	Arg R	CGT	26	0.76
	GTC	7	0.2		CGC	0	0
	GTA	80	2.34		CGA	4	0.12
	GTG	32	0.93		CGG	1	0.03
Ser (S1) (AGN)	TCT	136	3.97	Ser (S2) (UCN)	AGT	142	4.15
	TCC	4	0.12		AGC	3	0.09
	TCA	19	0.55		AGA	61	1.78
	TCG	3	0.09		AGG	23	0.67
Pro (P)	CTT	56	1.64	Thr (T)	ACT	89	0.26
	CCC	5	0.15		ACC	3	0.09
	CCA	16	0.47		ACA	21	0.61
	CCG	6	0.18		ACG	8	0.23
Ala (A)	GCT	83	2.42	Gly (G)	GGT	126	3.68
	GCC	4	0.12		GGC	6	0.18
	GCA	8	0.23		GGA	35	1.02
	GCG	4	0.12		GGG	22	0.64
His (H)	CAT	55	1.61	Gln (Q)	CAA	22	0.64
	CAC	2	0.06		CAG	20	0.58
Cyc C	TGT	51	1.49	Trp (W)	TGA	49	1.43
	TGC	0	0		TGG	23	0.67

A total of 3424 codons were analysed, excluding the initiation and termination codons.
* Termination codons.

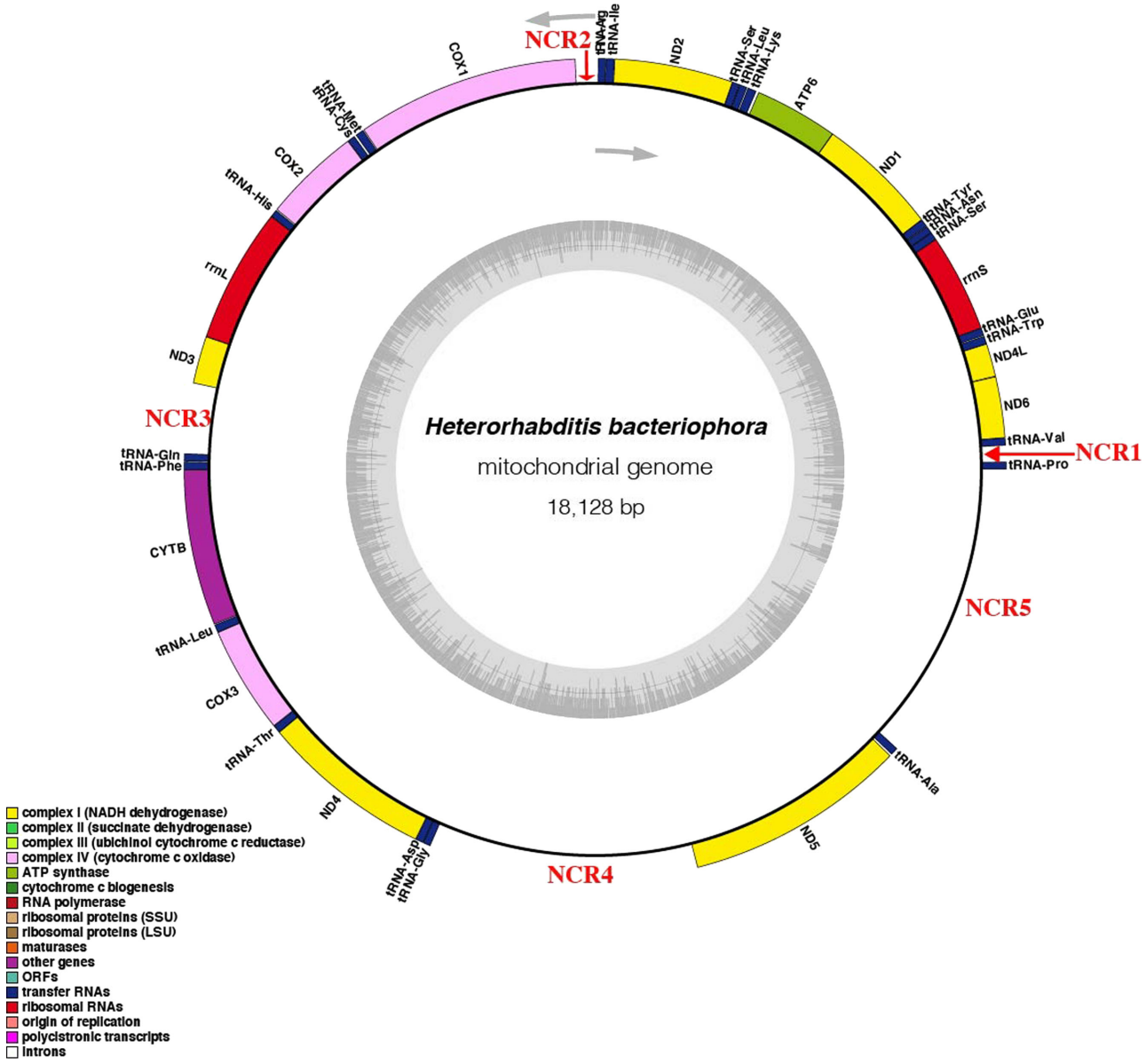


Fig. 1. Circular map of the mitochondrial genome of *Heterorhabditis bacteriophora*. The inner ring shows percentage GC content. Arrows indicate relative transcriptional orientation. NCRs 1-5 are labelled. Note: genes *ND1* to *ND6* on the map are written as *nad1* to *nad6* in the text and Table 1.

rRNA genes of *H. bacteriophora* were compared with those of a representative set of nematodes: *C. elegans*, *Ascaris suum*, *Ancylostoma duodenale*, *Necator americanus*, *Cooperia oncophora*, *Steinernema carpocapsae*, *Strongyloides stercoralis*, *Onchocerca volvulus*, *Dirofilaria immitis*, *Brugia malayi*, *T. spiralis* and *Xiphinema americanum*. Pairwise comparisons revealed sequence identities of 21-86% among the compared nematode species. The most conserved proteins were *cox1* and *cox2*, while *nad6* and *nad2* were the least conserved. For all compared proteins, the amino acid sequence identity was higher among nematodes of order Rhabditida (*e.g.*, *C. elegans* and *Cooperia oncophora*) than among members of the orders Spirurida (*e.g.*, *O. volvulus*) and Trichinellida (*T. spiralis*).

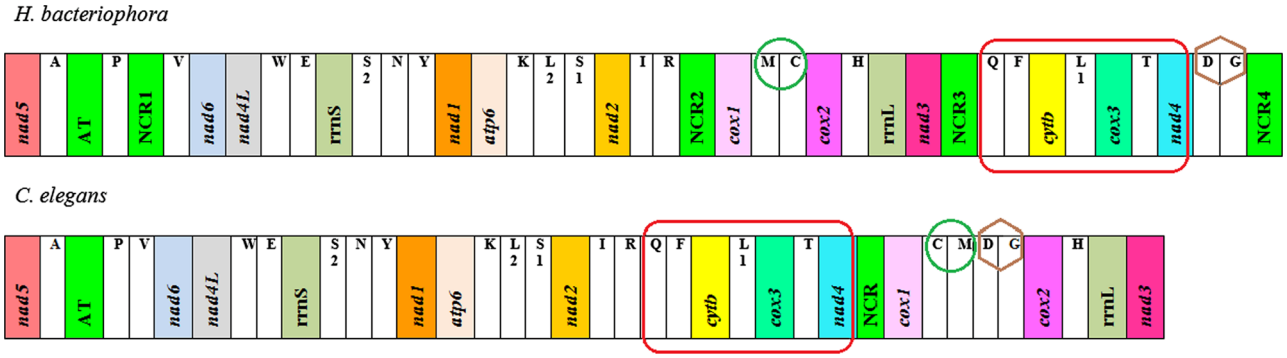


Fig. 2. Comparison of linearised representation of mitochondrial gene arrangement of *Heterorhabditis bacteriophora* and the free-living nematode *Caenorhabditis elegans*. The circular mitochondrial genomes were linearised at the 5′ end of *nad5*. Gene, NCR and genome size are not drawn to scale. The red rectangular box indicates a translocation of a block of seven genes, *trnQ-trnF-cytb-trnL1-cox3-trnT-nad4*, between *nad3* and *nad5*, as well as a change in the position of a block of four tRNAs, *trnC-trnM-trnD-trnG*, where *trnC* and *trnM* have switched places (green circle), and *trnD* and *trnG* (brown hexagon) have translocated between the *nad4* and *nad5* genes, relative to their position in *C. elegans*.

Transfer RNA and ribosomal RNA genes

The anticodon triplet sequences of the 22 encoded tRNA genes (Fig. 1) are identical to their counterparts in *C. elegans*, *A. suum* (Okimoto *et al*., 1992), *A. duodenale* and *N. americanus* (Hu *et al*., 2002). Twenty of these tRNAs lack the TψC loop and arm, a feature that is also found in other secernentian nematodes (Wolstenholme *et al*., 1987; Okimoto *et al*., 1992; Hu & Gasser, 2006). These tRNA genes are between 54 and 63 nucleotides in size, and their secondary structure is very similar to *A. duodenale* and *N. americanus* (Hu *et al*., 2002). Each tRNA predicted from these genes has an aminoacyl stem of 6-7 bp, a DHU stem of 4 bp (3 bp in tRNA A), a DHU loop of between 4-7 nucleotides, an anticodon stem of 4-5 bp, an anticodon loop of 7 nucleotides, and a TV replacement loop of 7-13 nucleotides.

The small and large subunit rRNAs (*rrnS* and *rrnL*) of *H. bacteriophora* are predicted to be 701 and 965 bp in size (Table 1) and their AT content is 71.9% and 78.03%, respectively. These lengths, as well as their locations, are similar to the length of the ribosomal RNAs of *C. elegans* and of other nematodes including *A. suum*, *A. duodenale*, *N. americanus*, *C. oncophora* and *S. carpocapse* (Okimoto *et al*., 1992; Hu *et al*., 2002; Van der Veer & de Vries, 2004; Montiel *et al*., 2006). However, needle (*X. americanum*, *X. rivesi*, *X. pachtaicum*) and dagger (*Longidorus vineacola*, *Paralongidorus litoralis*) enoplean nematodes have smaller rRNA genes; the *rrnS* range from 531 to 606 bp, while the *rrnL* range from 707 to 729 bp (Palomares-Rius *et al*., 2017).

Gene order

The gene order of the *H. bacteriophora* mitochondrial genome is distinct from that of all other nematodes sequenced to date. Its gene order differs from that of *C. elegans* in having a translocation of a block of seven genes, *trnQ-trnF-cytb-trnL1-cox3-trnT-nad4*, located between *nad3* and *nad5*, as well as a change in the position of a block of four tRNAs, *trnC-trnM-trnD-trnG*, where *trnC* and *trnM* have switched places, and *trnD* and *trnG* have translocated between the *nad4* and *nad5* genes, relative to their position in *C. elegans* (Figs 1, 2; Table 1). When translocations of tRNA genes are excluded, (which are more frequent than those of rRNA and protein coding genes; Wolstenholme, 1992; Boore, 1999; Gissi *et al*., 2008), *C. elegans* shares its gene order with 27 out of 28 species of Rhabditomorpha and nine out of 11 species of Ascaridomorpha (Kim *et al*., 2017). A survey of 65 nematode mitochondrial genomes by Liu *et al*. (2013) revealed 25 different nematode gene arrangement (GA) patterns. Further genome sampling has shown that the GA patterns of nematodes from the chromadorean infraorder Tylenchomorpha are also variable, adding a further six patterns to the list of described GA in nematodes (Sultana *et al*., 2013; García & Sánchez-Puerta, 2015). The seven gene block translocation, which is the major distinguishing feature of *H. bacteriophora* GA relative to that of *C. elegans*, is flanked on each side by non-coding DNA NCR3 and NCR4 (Fig. 2).

The most commonly invoked model of mitochondrial gene rearrangement is the tandem duplication-random

loss model (Moritz & Brown, 1987; Boore & Brown, 1998; Muller & Boore, 2005; Chong & Mueller, 2017), as a result of slipped strand mispairing during mtDNA replication, followed by multiple deletions of redundant genes. Alternative mechanisms include the illegitimate elongation model (Buroker *et al*., 1990) depicted by error in termination of replication and transposition (Macey *et al*., 1997), a mechanism involving terminal direct repeats and intramolecular recombination that involves proximity to an origin of replication (Lunt & Hyman, 1997) or a stem-loop structure (Stanton *et al*., 1994). The proposed mechanism of gene rearrangement in the mitochondrial genome of *H. bacteriophora* is the tandem duplication-random loss model, although other possible mechanisms such as intramolecular recombination cannot be ruled out.

Non-coding regions

Non-coding regions (NCR) account for 26.7% of the mitogenome and contribute to its relatively large size. The genome contains a total of 13 short intergenic sequences, ranging in size from 1 to 18 bp (Table 1). The positions and sizes of these short intergenic sequences are similar to those of *C. elegans* (Okimoto *et al*., 1992). The genome also has five non-coding regions, NCR1-NCR5 (Fig. 1; Table 1), whose sequences are longer than 18 bp and have no BLASTN or BLASTX similarities in GenBank. The majority of nematode mitogenomes contain two non-coding regions, the long (major) non-coding region (AT-rich region) and the short (minor) non-coding region. The major non-coding region, also called the control region (CR), contains elements involved in the regulation of replication and transcription (Wolstenholme, 1992; Boore, 1999). These include tandem repeat sequences and inverted repeats, which may form hairpin and loop structures that likely act as signals for polymerases (Brenton *et al*., 2014). Extensive variation in the size and sequence content of the control region occurs in many species of invertebrates, including nematodes (Zang & Hewitt, 1997; Sun *et al*., 2014; Li & Liang, 2018).

NCR5 is the largest non-coding region (2154 bp) and has the highest AT content (83.4%). Based on its AT content and location in the mtDNA molecule, NCR5 is considered to be the putative AT-rich control region. Its location between trnA and trnP corresponds to the location of the AT-rich regions of *C. elegans* (Okimoto *et al*., 1992), *A. duodenale* and *N. americanus* (Hu *et al*., 2002), *C. oncophora* (Van der Veer & de Vries, 2004) and *Haemonchus contortus* (Jex *et al*., 2008). However, NCR5 has no significant sequence similarity to any of the above-mentioned AT-rich regions. The repeated sequence motifs (CR1-CR6) present in the *C. elegans* AT-rich region were not found in NCR5, which also lacked any of the inverted repeat sequence tracts found in *A. suum*, *O. volvulus* and *A. duodenale*, or *N. americanus* (Okimto *et al*., 1992; Keddie *et al*., 1998; Hu *et al*., 2002). Instead, two (AT) dinucleotide repeat regions ($(AT)_{23}$ and $(AT)_{55}$) were found, which are similar to a run of 18 AT dinucleotide in *C. elegans* and *A. suum* (Okimoto *et al*., 1992). In addition, NCR5 also contained two (TA) dinucleotide repeat regions ($(TA)_{13}$ and $(TA)_{28}$).

NCR5 is relatively large, comprising 2154 bp; however, the size of the control region in nematodes is very variable (Supplementary Table S2). The shortest AT-rich regions occur in plant-parasitic nematodes of the order Dorylaimida: *Xiphinema* spp. (95-140 bp) and *L. vineacola* (92 bp) (Palomares-Rius *et al*., 2017), while the tylenchid plant parasites *Pratylenchus vulnus* and *Meloidogyne graminicola* possess large AT-rich regions of 6847 bp and 5063 bp, respectively (Sultana *et al*., 2013; Besnard *et al*., 2014). The insect-parasitic nematode, *S. litorale*, also has a large AT-rich region comprising 6260 bp (Kikuchi *et al*., 2016), but, interestingly, the AT-rich regions of the congeneric *S. carpocapsae* and *S. kushidai* are 554 and 643 bp, respectively (Montiel *et al*., 2006; Kikuchi *et al*., 2016), an indication of the extent of length variation that can be found in the AT-rich region within a single genus of nematodes. NCR2 most likely corresponds to the second (minor) non-coding region found in *C. elegans* and *A. suum* (Okimoto *et al*., 1992), *A. duodenale* and *N. americanus* (Hu *et al*., 2002), *C. oncophora* (Van der Veer & de Vries, 2004), *D. immitis* (Hu *et al*., 2003) and *S. carpocapsae* (Montiel *et al*., 2006). It contains a poly-A (7 nt) region, which is also found in the minor non-coding regions of *C. elegans*, *A. suum*, *A. duodenale*, *N. americanus* and *D. immitis*. Hu *et al*. (2002) suggested that probably this non-coding region is the site for the initiation of the second strand synthesis, although there is no experimental evidence for this.

NCR1 (114 bp) is located between *trnP* and *trnV* (Table 1). Its nucleotide sequence had little homology to the other four NCR regions (Fig. 3). Its AT content is 34.7%; a stem loop structure could be predicted for this region. NCR3 (498 bp) is located between *nad3* and *trnQ*. The complete 498 bp NCR3 sequence is duplicated in NCR4 and in NCR5 (the putative AT-rich control region) and it also has sequence similarity to 79 nucleotides at the 3′ end of NCR2, the putative minor control region. NCR4 (1941 bp) is located between *trnG* and *nad5*. At

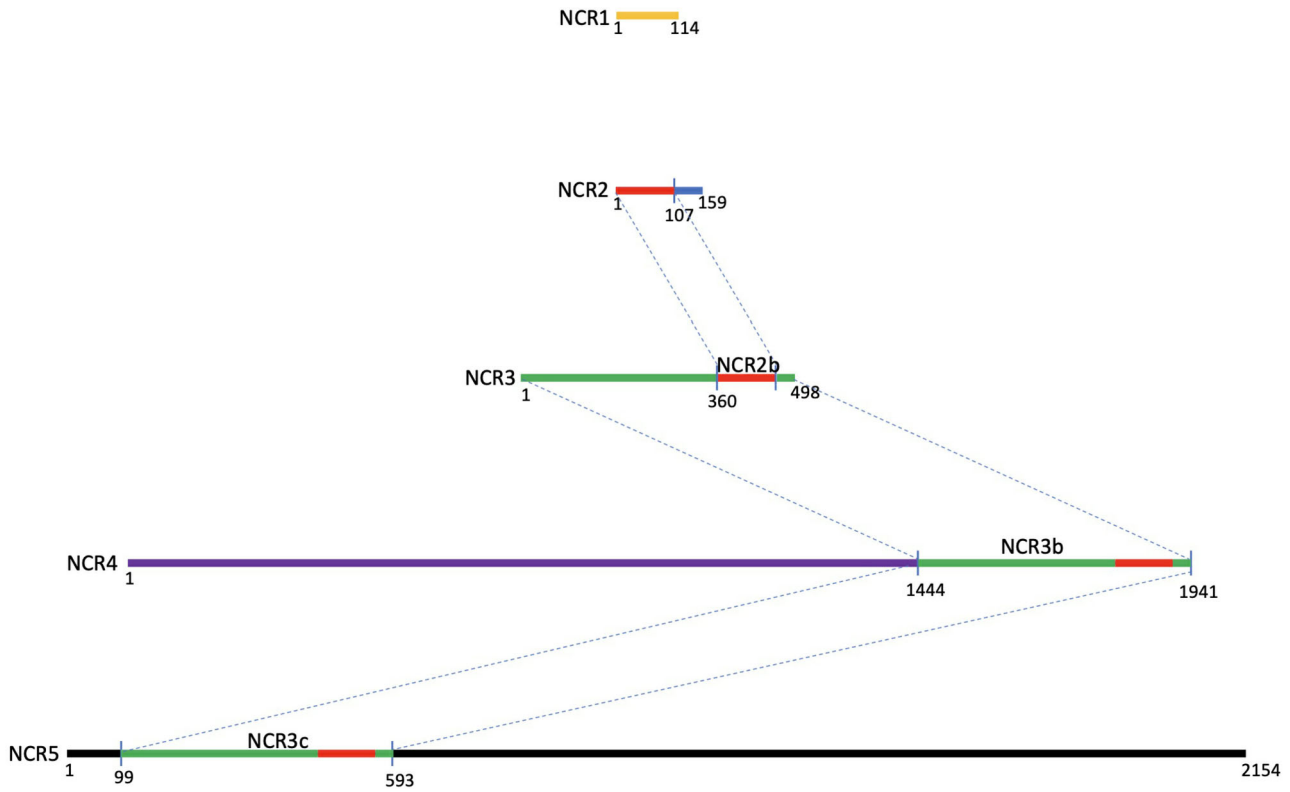


Fig. 3. NCRs of *Heterohabditis bacteriophora*. Homologous regions of repeated DNA together with their relative positions are colour coded and shown. NCRs 2-5 share segments of sequence with homology resulting from duplication. NCR1 shares no homology with NCR 2-4.

the 3′ end of NCR4, the NCR3 repeat sequence overlaps with the first 24 nucleotides of the *nad5* gene. The three full-length copies of the NCR3 sequence that share 100% nucleotide sequence identity have been designated as NCR3a, NCR3b and NCR3c (Fig. 1; Table 1). Such an organisation has not been reported previously in nematode mtDNA.

A single common motif of 30 bp: AGAAAGGAGGAGGCAGGAGCTCCCCTTTCT, was found in all five NCR sequences using both the SeSiMCMC program (Favorov *et al*., 2005) and the MEME-motif discovery tool (Baily & Elkan, 1994). This motif has no counterpart in other nematodes examined so far. It can form a hairpin structure, but the loop does not contain a run of Ts. The motif was also found in NCR1 on the opposite strand, with a few mismatch pairings, and it is also predicted to form a stable hairpin structure. Within this 30 bp motif there is a smaller 5 bp motif: TCCCC. There are two copies of the TCCCC. It is also present twice in NCR5 (*i.e.*, AT-rich control region) outside of the repeated copy of NCR3c. A TCCCC motif has been found in the D-Loop of some mammals (*e.g.*, pigs and cattle) (Douzery & Randi, 1997) and some birds (Randi & Lucchini, 1998; Eberhard *et al*., 2001), and this motif has been functionally associated with termination of nascent H strand during mtDNA replication (Dufresne *et al*., 1996; Ritchie & Lambert, 2000; Pie *et al*., 2008; Wang *et al*., 2011; Shi *et al*., 2012); whether this motif has a similar function in the *H. bacteriophora* mtDNA replication needs experimental evidence.

The presence of three copies of the 498 bp NCR3 sequence in the *H. bacteriophora* mitogenome most probably confers a selective advantage, since the mode of evolution in metazoan mtDNA has been selection for small genome size (Attardi, 1985), and non-functional sequences would be expected to be eliminated relatively quickly due to the more rapid rate of replication of compact mitogenomes (Schirtzinger *et al*., 2012; Kinkar *et al*., 2019). Moreover, an assay for the presence of the NCR3 by PCR and DNA sequencing confirmed its existence in seven other *Heterorhabditis* species: *H. megidis* (MW512820), *H. zealandica* (MW512821), *H. marela-*

tus (MW512822), *H. downesi* (MW512823), *H. taysearae* (MW512824), *H. mexicana* (MW512825) and *H. indica* (MW512826), with nearly 100% nucleotide sequence identity with *H. bacteriophora* (MW512819) (Supplementary Fig. S2), although the number of repeated copies was not determined. This suggests concerted evolution of NCR3 region among *Heterorhabditis* species.

In total, there are five NCR in the *H. bacteriophora* mitochondrial genome (Fig. 1). Sequence analysis indicates that there is sequence homology between different NCR (Fig. 3). For example, the first 107 nucleotides of NCR2 overlap with positions 360-498 of NCR3 (99% identity). NCR3 itself overlaps with positions 1444-1941 of NCR4 (99% identity) and positions 99-593 (99% identity) of NCR5. There are two equally plausible scenarios that may have led to this NCR sequence arrangement. The first involves the duplication of NCR3, three times with subsequent rearrangement/insertion to NCR4 and NCR5 regions. The third duplicated copy of NCR3 was then cleaved, resulting in NCR2. Alternatively, NCR2 may be the progenitor copy that duplicated and inserted into NCR3a, followed by two subsequent duplications of NCR3a and insertion into NCR4 and NCR5 regions.

Duplications of the control regions of mtDNA molecules have been described in diverse species of vertebrates such as fish (Inoue *et al*., 2003; Tatarenkov & Avise 2007), reptiles (Kumazawa *et al*., 1996; Quian *et al*., 2018) and birds (Schirtzinger *et al*., 2012), as well as in ostracods (Ogoh & Ohmiya, 2007), insects (Shao & Barker, 2003; Li *et al*., 2017), arachnids (Shao *et al*., 2005), squids (Tomita *et al*., 2002; Jiang *et al*., 2018) and the cestode *Echinococcus granulosus* (Kinkar *et al*., 2019). Three nearly identical copies of a *ca* 500 bp of non-coding region have been reported in the squid *Loligo bleekeri*, for which several stem and loop structures could be predicted (Tomita *et al*., 2002). These authors suggested these secondary structures may play a role in the mtDNA replication or transcription processes. Recently Jiang *et al*. (2018) found that eight of the nine squid species investigated share the same genome organisation as *L. bleekeri* and also possess three dispersed putative control regions. In a study of 114 species of parrots, Schirtzinger *et al*. (2012) found that 76 species had a single mtDNA control region as is typical of birds generally, but a duplication of the mtDNA control region had occurred in 38 parrot species. Phylogenetic analysis identified at least six independent origins of these control region duplications and, interestingly, there were no reversions to a single control region state in any of these six clades. The authors postulate that the advantage of having a second control region overrides selection for compactness, possibly by allowing for faster replication of the mtDNA molecule (Schirtzinger *et al*., 2012). Similarly, the presence of two additional dispersed copies of the control region NC3 sequence may perhaps confer a replicative advantage to *H. bacteriophora* mtDNA, particularly during the phase of rapid growth and reproduction within the insect cadaver.

Conclusions

We describe here the mtDNA genome of *H. bacteriophora*, the first description of a complete mtDNA sequence for a member of the family Heterorhabditidae. This genome contains the typical chromadorean complement of 12 protein-coding genes, 22 tRNA genes and two rRNA genes; however, it is 18 128 bp long and is thus larger than most chromadorean mitogenomes, which typically range from 13 to 15 kb. The relatively large size of the *H. bacteriophora* mitogenome results from the presence of five non-coding regions (NCR1-NCR5), which make up 26.7% of the genome. NCR5, the largest non-coding region (2154 bp), has the highest AT content and is most likely the control region. The complete 498 bp sequence of NCR3 also occurs in the NCR5 (control region) sequence and the NCR4 sequence. The occurrence of three conserved copies of a relatively large dispersed repeat sequence has not been previously observed in nematodes, although duplications of mtDNA control regions have been described in diverse species of vertebrates and invertebrates. It may be possible that the three copies of the 498 bp NCR3 sequence correspond to three dispersed putative control regions that have been retained because they confer a replicative advantage to the *H. bacteriophora* mtDNA molecule. *Heterorhabditis bacteriophora* has an idiosyncratic mtDNA gene order that differs from that of *C. elegans* in having a translocation of a block of seven genes located between *nad3* and *nad5*, as well as a change in the position of a block of four tRNAs. This *H. bacteriophora* genome will provide a useful reference resource for designing primers to conduct population genetic studies and for investigating and understanding the molecular and evolutionary biology of nematode mitochondrial genomes. Further research is needed to address the functional significance of the non-coding regions, the mechanisms by which the NCR3 sequence has been triplicated in the mitochondrial genome of *H. bacteriophora* and has been conserved in the mitogenomes of other *Heterorhabditis* species.

Authors' note

While this manuscript was in preparation, the sequence of the mtDNA genome of *Heterorhabditis indica* was submitted to GenBank under the Accession number NC_040293.

Acknowledgement

This study was funded by the Libyan Ministry of Higher Education and the University of Tripoli (previously known as Al-Fatah University), Faculty of Science, Tripoli, Libya.

References

Ahmad, A.A., Yang, X., Zhang, T., Wang, C., Zhou, C., Yan, X., Hassan, M., Ikram, M. & Hu, M. (2019). Characterization of the complete mitochondrial genome of *Ostertagia trifurcata* of small ruminants and its phylogenetic associations for the Trichostrongyloidea superfamily. *Genes* 10, 107. DOI: 10.3390/genes10020107

Akhurst, R.J. & Bedding, R.A. (1986). Natural occurrence of insect pathogenic nematodes (Steinernematidae and Heterorhabditidae) in soil in Australia. *Journal of the Australian Entomology Society* 25, 241-244. DOI: 10.1111/j.1440-6055.1986.tb01110.x

Albu, M., Min, X.J., Hickey, D. & Golding, B. (2008). Uncorrected nucleotide bias in mtDNA can mimic the effects of positive Darwinian selection. *Molecular Biology and Evolution* 25, 2521-2524. DOI: 10.1093/molbev/msn224

Attardi, G. (1985). Animal mitochondrial DNA: an extreme example of genetic economy. *International Review of Cytology* 93, 93-145. DOI: 10.1016/S0074-7696(08)61373-X

Avise, J.C. (1986). Mitochondrial DNA and the evolutionary genetics of higher animals. *Philosophical Transactions of the Royal Society B* 312, 325-342. DOI: 10.1098/rstb.1986.0011

Bai, X., Adams, B.J., Ciche, T.A., Clifton, S., Gaugler, R., Kim, K.W., Spieth, J., Sternberg, P.W., Wilson, R.K. & Grewal, P.S. (2013). A lover and a fighter: the genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. *PLOS ONE* 8, e69618. DOI: 10.1371/journal.pone.0069618

Bailey, T. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymer. *Proceedings/International Conference on Intelligent Systems for Molecular Biology* 2, 28-36.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27, 573-580. DOI: 10.1093/nar/27.2.573

Besnard, G., Jühling, F., Chapuis, E., Zedane, L., Lhuillier, E., Mateille, T. & Bellafiore, S. (2014). Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. *Comptes Rendus Biologies* 337, 295-301. DOI: 10.1016/j.crvi.2014.03.003

Bhat, A.H., Chaubey, A.K. & Askary, T.H. (2020). Global distribution of entomopathogenic nematodes, *Steinernema* and *Heterorhabditis*. *Egyptian Journal of Biological Pest Control* 30, 31. DOI: 10.1186/s41938-020-0212-y

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. *et al.* (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature.* 392, 71-75. DOI: 10.1038/32160

Blouin, M.S. (1998). Mitochondrial DNA diversity in nematodes. *Journal of Helminthology* 72, 285-289. DOI: 10.1017/s0022149x00016618

Blouin, M.S. (2002). Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. *International Journal for Parasitology* 32, 527-531. DOI: 10.1016/s0020-7519(01)00357-5

Blouin, M.S., Yowell, C.A., Courtney, C.H. & Dame, J.B. (1998). Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Molecular Biology and Evolution* 15, 1719-1727. DOI: 10.1093/oxfordjournals.molbev.a025898

Boore, J.L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research* 27, 1767-1780. DOI: 10.1093/nar/27.8.1767

Boore, J.L. & Brown, W.M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics & Development* 8, 668-674. DOI: 10.1016/s0959-437x(98)80035-x

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T. & Passamonti, M.A. (2014). A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends in Genetics* 30, 555-564. DOI: 10.1016/j.tig.2014.09.002

Burnell, A.M. & Stock, S.P. (2000). *Heterorhabditis*, *Steinernema* and their bacterial symbionts – lethal pathogens of insects. *Nematology* 2, 31-42. DOI: 10.1163/156854100508872

Buroker, N.E., Brown, J.R., Gilbert, T.A., O'Hara, P.J., Beckenbach, A.T., Thomas, W.K. & Smith, M.J. (1990). Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* 124, 157-163.

Chong, R.A. & Mueller, R.L. (2017). Polymorphic duplicate genes and persistent non-coding sequences reveal heterogeneous patterns of mitochondrial DNA loss in salamanders. *BMC Genomics* 18, 992. DOI: 10.1186/s12864-017-4358-2

Derycke, S., Backeljau, T. & Moens, T. (2013). Dispersal and gene flow in free-living marine nematodes. *Frontiers in Zoology* 10, 1. DOI: 10.1186/1742-9994-10-1

DeSalle, R., Schierwater, B. & Hadrys, H. (2017). MtDNA: the small workhorse of evolutionary studies. *Frontiers in Bioscience* 22, 873-887. DOI: 10.2741/4522

Dolan, K.M. (2001). *The genetic enhancement and analysis of dauer juvenile recovery in the entomopathogenic nematode* Heterorhabditis bacteriophora. Ph.D. Thesis. National University of Ireland, Maynooth, Ireland.

Douzery, E. & Randi, E. (1997). The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Molecular Biology and Evolution* 14, 1154-1166. DOI: 10.1093/oxfordjournals.molbev.a025725

Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.-F. *et al.* (2003). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nature Biotechnology* 21, 1307-1313. DOI: 10.1038/nbt886

Dufresne, C., Mignote, F. & Gueride, M. (1996). The presence of tandem repeats and the initiation of replication in rabbit mitochondrial DNA. *European Journal of Biochemistry* 235, 593-600. DOI: 10.1111/j.1432-1033.1996.00593.x

Dunphy, G.B. & Webster, J.M. (1989). The monoxenic culture of *Neoaplectana carpocapsae* DD 136 and *Heterorhabditis heliothidis*. *Revue de Nématologie* 12, 113-123.

Eberhard, J.R., Wright, T.M. & Bermingham, E. (2001). Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. *Molecular Biology and Evolution* 18, 1330-1342. DOI: 10.1093/oxfordjournals.molbev.a003917

Ehlers, R.-U. (2001). Mass production of entomopathogenic nematodes for plant protection. *Applied Microbiology and Biotechnology* 56, 623-633. DOI: 10.1007/s002530100711

Favorov, A.V., Gelfand, M.S., Gerasimova, A.V., Ravcheev, D.A., Mironov, A.A. & Makeev, V.J. (2005). A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length. *Bioinformatics* 21, 2240-2245. DOI: 10.1093/bioinformatics/bti336

García, L.E. & Sánchez-Puerta, M.V. (2015). Comparative and evolutionary analyses of *Meloidogyne* spp. based on mitochondrial genome sequences. *PLoS ONE* 10, e0121142. DOI: 10.1371/journal.pone.0121142

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.D., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D. *et al.* (2007). Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, 1756-1760. DOI: 10.1126/science.1145406

Gissi, C., Iannelli, F. & Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 101, 301-320. DOI: 10.1038/hdy.2008.62

Greiner, S., Lehwark, P. & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47(W1), W59-W64. DOI: 10.1093/nar/gkz238

Harrison, R.G. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution* 4, 6-11. DOI: 10.1016/0169-5347(89)90006-2

He, Y., Jones, J., Armstrong, M., Lamberti, F. & Moens, M. (2005). The mitochondrial genome of *Xiphinema americanum sensu stricto* (Nematoda: Enoplea): considerable economization in the length and structural features of encoded genes. *Journal of Molecular Evolution* 61, 819-833. DOI: 10.1007/s00239-005-0102-7

Hu, M. & Gasser, R.B. (2006). Mitochondrial genomes of parasitic nematodes – progress and perspectives. *Trends in Parasitology* 22, 78-84. DOI: 10.1016/j.pt.2005.12.003

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. DOI: 10.1016/s0020-7519(01)00316-2

Hu, M., Chilton, N.B. & Gasser, R.B. (2003a). The mitochondrial genome of *Strongyloides stercoralis* (Nematoda)-idiosyncratic gene order and evolutionary implications. *International Journal for Parasitology* 33, 1393-1408. DOI: 10.1016/s0020-7519(03)00130-9

Hu, M., Gasser, R.B., Abs El-Osta, Y.G. & Chilton, N.B. (2003b). Structure and organization of the mitochondrial genome of the canine heartworm, *Dirofilaria immitis*. *Parasitology* 127, 37-51. DOI: 10.1017/S0031182003003275

Humphreys-Pereira, D.A. & Elling, A.A. (2015). Mitochondrial genome plasticity among species of the nematode genus *Meloidogyne* (Nematoda: Tylenchina). *Gene* 560, 173-183. DOI: 10.1016/j.gene.2015.01.065

Inoue, J.G., Miya, M., Tsukamoto, K. & Nishida, M. (2003). Evolution of the deep-sea gulper eel mitochondrial genomes: large-scale gene rearrangements originated within the eels. *Molecular Biology and Evolution* 20, 1917-1924. DOI: 10.1093/molbev/msg206

Jex, A.R., Hu, M., Timothy, D., Littlewood, J., Waeschenbach, A. & Gasser, R.B. (2008). Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). *BMC Genomics* 9, 11. DOI: 10.1186/1471-2164-9-11

Jiang, L., Kang, L., Wu, C., Chen, M. & Lü, Z. (2018). A comprehensive description and evolutionary analysis of 9 Loliginidae mitochondrial genomes. *Hydrobiologia* 808, 115-124. DOI: 10.1007/s10750-017-3377-y

Joyce, S.A., Burnell, A.M. & Powers, T.O. (1994). Characterization of *Heterorhabditis* isolates by PCR amplification of segments of mtDNA and rDNA genes. *Journal of Nematology* 26, 260-270.

Kang, S., Sultana, T., Eom, K.S., Park, Y.C., Soonthornpong, N., Nadler, S.A. & Park, J.K. (2009). The mitochondrial genome sequence of *Enterobius vermicularis* (Nematoda: Oxyurida) – an idiosyncratic gene order and phylogenetic information for chromadorean nematodes. *Gene* 429, 87-97. DOI: 10.1016/j.gene.2008.09.011

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. DOI: 10.1016/s0166-6851(98)00102-9

Kikuchi, T., Afrin, T. & Yoshida, M. (2016). Complete mitochondrial genomes of four entomopathogenic nematode species of the genus *Steinernema*. *Parasites & Vectors* 9, 430. DOI: 10.1186/s13071-016-1730-z

Kim, J., Kern, E., Kim, T., Sim, M., Kim, J., Kim, Y., Park, C., Nadler, S.A. & Park, J.K. (2017). Phylogenetic analysis of two *Plectus* mitochondrial genomes (Nematoda: Plectida) supports a sister group relationship between Plectida and Rhabditida within Chromadorea. *Molecular Phylogenetics and Evolution* 107, 90-102. DOI: 10.1016/j.ympev.2016.10.010

Kim, K.-H., Eom, K.S. & Park, J.-K. (2006). The complete mitochondrial genome of *Anisakis simplex* (Ascaridida: Nematoda) and phylogenetic implications. *International Journal for Parasitology* 36, 319-328. DOI: 10.1016/j.ijpara.2005.10.004

Kinkar, L., Korhonen, P.K., Cai, H., Gauci, C.G., Lightowlers, M.W., Saarma, U., Jenkins, D.J., Li, J., Li, J., Young, N.D. *et al.* (2019). Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1. *Parasites & Vectors* 12, 238. DOI: 10.1186/s13071-019-3492-x

Kumazawa, Y., Ota, H., Nishida, M. & Ozawa, T. (1996). Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular Biology and Evolution* 13, 1242-1254. DOI: 10.1093/oxfordjournals.molbev.a025690

Labaude, S. & Griffin, C.T. (2018). Transmission success of entomopathogenic nematodes used in pest control. *Insects* 9, 72. DOI: 10.3390/insects9020072

Lacey, L.A. & Georgis, R. (2012). Entomopathogenic nematodes for control of insect pests above and below ground with comments on commercial production. *Journal of Nematology* 44, 218-225.

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. PMCID: PMC1461501

Li, H., Song, F., Gu, W., Feng, J., Cai, W. & Shao, R. (2017). Novel insights into mitochondrial gene rearrangement in thrips (Insecta: Thysanoptera) from the grass thrips, *Anaphothrips obscurus*. *Scientific Reports* 7, 4284. DOI: 10.1038/s41598-017-04617-5

Li, K. & Liang, A.-P. (2018). Hemiptera mitochondrial control region: new sights into the structural organization, phylogenetic utility, and roles of tandem repetitions of the noncoding segment. *International Journal of Molecular Sciences* 19, 1292. DOI: 10.3390/ijms19051292

Liu, G.-H., Wang, Y., Xu, M.-J., Zhou, D.-H., Ye, Y.-G., Li, J.-Y., Song, H.-Q., Lin, R.-Q. & Zhu, X.-Q. (2012). Characterization of the complete mitochondrial genomes of two whipworms *Trichuris ovis* and *Trichuris discolor* (Nematoda: Trichuridae). *Infection Genetics Evolution* 12, 1635-1641. DOI: 10.1016/j.meegid.2012.08.005

Liu, G.-H., Shao, R., Li, J.-Y., Zhou, D.-H., Li, H. & Zhu, X.-Q. (2013). The complete mitochondrial genomes of three parasitic nematodes of birds: a unique gene order and insights into nematode phylogeny. *BMC Genomics* 14, 414. DOI: 10.1186/1471-2164-14-414

Liu, G.-H., Li, S., Zou, F.-C., Wang, C.-R. & Zhu, X.-Q. (2016). The complete mitochondrial genome of rabbit pinworm *Passalurus ambiguus*: genome characterization and phylogenetic analysis. *Parasitolology Research* 115, 423-429. DOI: 10.1007/s00436-015-4778-3

Liu, J., Berry, R.E. & Blouin, M.S. (1999). Molecular differentiation and phylogeny of entomopathogenic nematodes (Rhabditida: Heterorhabditidae) based on ND4 gene sequences of mitochondrial DNA. *Journal of Parasitology* 85, 709-715. DOI: 10.2307/3285747

Lowe, T.M. & Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25, 955-964. DOI: 10.1093/nar/25.5.955

Lunt, D.H. & Hyman, B.C. (1997). Animal mitochondrial DNA recombination. *Nature* 387, 247. DOI: 10.1038/387247a0

Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z. & Papenfuss, T.J. (1997). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution* 14, 91-104. DOI: 10.1093/oxfordjournals.molbev.a025706

McLean, F., Berger, D., Laetsch, D.R., Schwartz, H.T. & Blaxter, M. (2018). Improving the annotation of the *Heterorhabditis bacteriophora* genome. *GigaScience* 7, giy034. DOI: 10.1093/gigascience/giy034

Mohandas, N., Pozio, E., La Rosa, G., Korhonen, P.K., Young, N.D., Koehler, A.V., Hall, R.S., Sternberg, P.W., Boag, P.R., Jex, A.R. *et al.* (2014). Mitochondrial genomes of *Trichinella* species and genotypes – a basis for diagnosis, and systematic and epidemiological explorations. *International Journal for Parasitology* 44, 1073-1080. DOI: 10.1016/j.ijpara.2014.08.010

Montiel, R., Lucena, M.A., Medeiros, J. & Simões, N. (2006). The complete mitochondrial genome of the entomopathogenic nematode *Steinernema carpocapsae*: insights into nematode mitochondrial DNA evolution and phylogeny. *Journal of Molecular Evolution* 62, 211-225. DOI: 10.1007/s00239-005-0072-9

Moritz, C. & Brown, W.M. (1987). Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proceedings of the National Academy*

of Sciences of the United States of America 84, 7183-7187. DOI: 10.1073/pnas.84.20.7183

Moritz, C., Dowling, T.E. & Brown, W.M. (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* 18, 269-292. DOI: 10.1146/annurev.es.18.110187.001413

Mueller, R.L. & Boore, J.L. (2005). Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Molecular Biology and Evolution* 22, 2104-2112. DOI: 10.1093/molbev/msi204

Nugent, M.J., O'Leary, S.A. & Burnell, A.M. (1996). Optimised procedures for the cryopreservation of different species of *Heterorhabditis*. *Fundamental and Applied Nematology* 19, 1-6.

Ogoh, K. & Ohmiya, Y. (2007). Concerted evolution of duplicated control regions within an ostracod mitochondrial genome. *Molecular Biology and Evolution* 24, 74-78. DOI: 10.1093/molbev/msl132

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.

Palomares-Rius, J.E., Cantalapiedra-Navarrete, C., Archidona-Yuste, A., Blok, V.C. & Castillo, P. (2017). Mitochondrial genome diversity in dagger and needle nematodes (Nematoda: Longidoridae). *Scientific Reports* 7, 41813. DOI: 10.1038/srep41813

Park, J.-K., Sultana, T., Lee, S.-H., Kang, S., Kim, H.L., Min, G.-S., Eom, K.S. & Nadler, S.A. (2011). Monophyly of clade III nematodes is not supported by phylogenetic analysis of complete mitochondrial genome sequences. *BMC Genomics* 12, 392. DOI: 10.1186/1471-2164-12-392

Pie, M.R., Oliveira-Neto, J.F., Boeger, W.A., Ostrensky, A. & Baggio, R.A. (2008). The organization of the mitochondrial control region in 2 brachyuran crustaceans: *Ucides cordatus* (Ocypodidae) and *Cardisoma guanhumi* (Gecarcinidae). *Journal of Heredity* 99, 432-437. DOI: 10.1093/jhered/esn024

Poinar Jr, G.O. (1976). Description and biology of a new insect parasitic rhabditoid, *Heterorhabditis bacteriophora* n. gen., n. sp. (Rhabditida; Heterorhabditidae n. fam.). *Nematologica* 21, 463-470. DOI: 10.1163/187529275X00239

Poinar Jr, G.O. (1979). *Nematodes for biological control of insects*. Boca Raton, FL, USA, CRC Press.

Poinar Jr, G.O. & Georgis, R. (1990). Characterization and field application of *Heterorhabditis bacteriophori* strain HP88 (Heterorhabditidae: Rhabditida). *Revue de Nématologie* 13, 387-393.

Powers, T.O., Harris, T.S. & Hyman, B.C. (1993). Mitochondrial DNA sequence divergence among *Meloidogyne incognita*, *Romanomermis culicivorax*, *Ascaris suum*, and *Caenorhabditis elegans*. *Journal of Nematology* 25, 564-572.

Quian, L., Wang, H., Yan, J., Pan, T., Jiang, S., Rao, D. & Zhang, B. (2018). Multiple independent structural dynamic events in the evolution of snake mitochondrial genomes. *BMC Genomics* 19, 354. DOI: 10.1186/s12864-018-4717-7

Randi, E. & Lucchini, V. (1998). Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. *Journal of Molecular Evolution* 47, 449-462. DOI: 10.1007/PL00006402

Ritchie, P.A. & Lambert, D.M. (2000). A repeat complex in the mitochondrial control region of Adélie penguins from Antarctica. *Genome* 43, 613-618. DOI: 10.1139/g00-018

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*, 2nd edition. New York, NY, USA, Cold Spring Harbor Laboratory Press.

Schirtzinger, E.E., Tavares, E.S., Gonzales, L.A., Eberhard, J.R., Miyaki, C.Y., Sanchez, J.J., Hernandez, A., Müeller, H., Graves, G.R., Fleischer, R.C. *et al.* (2012). Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Molecular Phylogenetics and Evolution* 64, 342-356. DOI: 10.1016/j.ympev.2012.04.009

Shao, R. & Barker, S.C. (2003). The highly rearranged mitochondrial genome of the plague thrips, *Thrips imagines* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Molecular Biology and Evolution* 20, 362-370. DOI: 10.1093/molbev/msg045

Shao, R., Barker, S.C., Mitani, H., Aoki, Y. & Fukunaga, M. (2005). Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. *Molecular Biology and Evolution* 22, 620-629. DOI: 10.1093/molbev/msi047

Sharp, P.M. & Matassi, G. (1994). Codon usage and genome evolution. *Current Opinion in Genetic and Development* 4, 851-860. DOI: 10.1016/0959-437X(94)90070-1

Shi, Y., Dierckx, A., Wanrooij, P.H., Wanrooij, S., Larsson, N.G., Wilhelmsson, L.M., Falkenberg, M. & Gustafsson, C.M. (2012). Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proceedings of the National Academy of Sciences of the United States of America* 109, 16510-16515. DOI: 10.1073/pnas.1119738109

Stanton, D.J., Daehler, L.L., Moritz, C. & Brown, W.M. (1994). Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* 137, 233-241.

Sultana, T., Kim, J., Lee, S.-H., Han, H., Kim, S., Min, G.-S., Nadler, S.A. & Park, J.-K. (2013). Comparative analysis of complete mitochondrial genome sequences confirms independent origins of plant-parasitic nematodes. *BMC Evolutionary Biology* 13, 1471-2148. DOI: 10.1186/1471-2148-13-12

Sun, L., Zhuo, K., Lin, B., Wang, H. & Liao, J. (2014). The complete mitochondrial genome of *Meloidogyne graminicola* (Tylenchina): a unique gene arrangement and its phylogenetic implications. *PLoS ONE* 9, e98558. DOI: 10.1371/journal.pone.0098558

Tang, S. & Hyman, B.C. (2007). Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei*. *Genetics* 176, 1139-1150. DOI: 10.1534/genetics.106.069518

Tatarenkov, A. & Avise, J.C. (2007). Rapid concerted evolution in animal mitochondrial DNA. *Proceedings of the Royal Society B* 274, 1795-1798. DOI: 10.1098/rspb.2007.0169

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. DOI: 10.1093/nar/25.24.4876

Tomita, K., Yokobori, S., Oshima, T., Ueda, T. & Watanabe, K. (2002). The cephalopod *Loligo bleekeri* mitochondrial genome: multiplied noncoding regions and transposition of tRNA genes. *Journal of Molecular Evolution* 54, 486-500. DOI: 10.1007/s00239-001-0039-4

Van der Veer, M. & de Vries, E. (2004). A single nucleotide polymorphism map of the mitochondrial genome of the parasitic nematode *Cooperia oncophora*. *Parasitology* 128, 421-431. DOI: 10.1017/s0031182003004633

Wang, L., Zhou, X. & Nie, L. (2011). Organization and variation of mitochondrial DNA control region in pleurodiran turtles. *Zoologia* 28, 495-504. DOI: 10.1590/S1984-46702011000400011

Wexler, Y., Yakhini, Z., Kashi, Y. & Geiger, D. (2005). Finding approximate tandem repeats in genomic sequences. *Journal of Computational Biology* 12, 928-942. DOI: 10.1089/cmb.2005.12.928

White, G.F. (1927). A method for obtaining infective larvae from cultures. *Science* 66, 302-303. DOI: 10.1126/science.66.1709.302-a

Wolstenholme, D.R. (1992). Animal mitochondrial DNA: structure and evolution. *International Review of Cytology* 141, 173-216. DOI: 10.1016/s0074-7696(08)62066-5

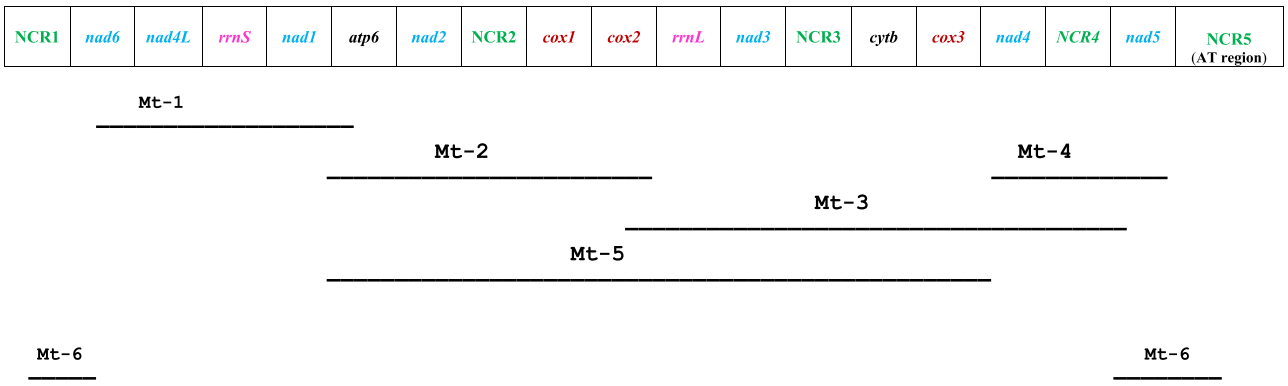
Wolstenholme, D.R., Macfarlane, J.L., Okimoto, R., Clary, D.O. & Whaleithner, J.A. (1987). Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proceedings of the National Academy of Sciences of the United States of America* 84, 1324-1328. DOI: 10.1073/pnas.84.5.1324

Wyman, S.K., Jansen, R.K. & Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 4, 1-4. DOI: 10.1093/bioinformatics/bth352

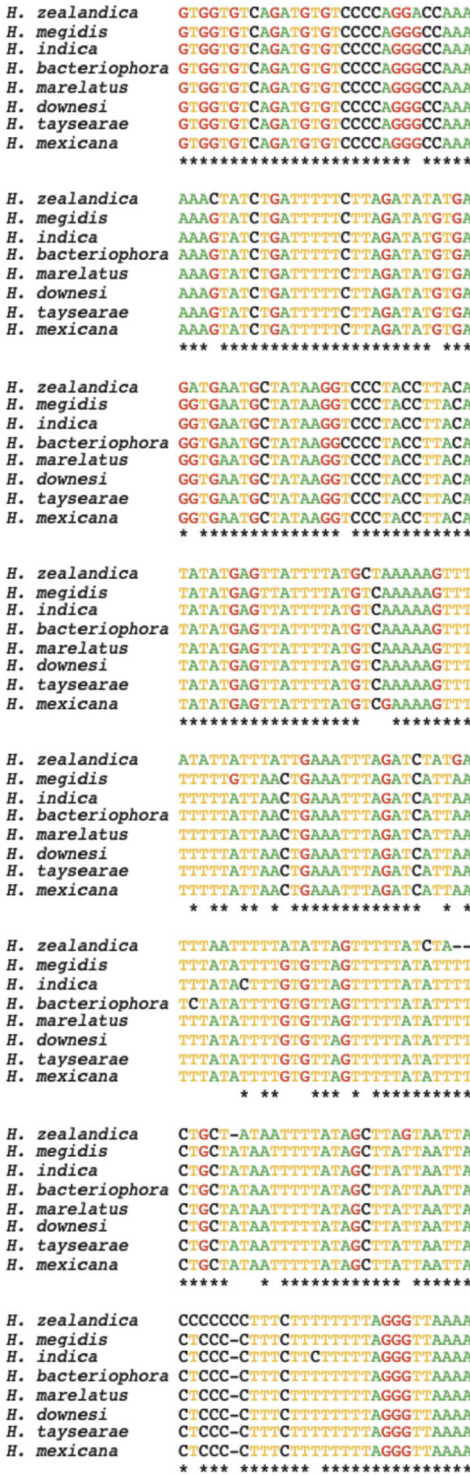
Zhang, D.-X. & Hewitt, G.M. (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* 25, 99-120. DOI: 10.1016/S0305-1978(96)00042-7

Zou, H., Jakovlić, I., Chen, R., Zhang, D., Zhang, J., Li, W.-X. & Wang, G.-T. (2017). The complete mitochondrial genome of parasitic nematode *Camallanus cotti*: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class. *BMC Genomics* 18, 840. DOI: 10.1186/s12864-017-4237-x

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31, 3406-3415. DOI: 10.1093/nar/gkg595



Supplementary Fig. S1. Amplification strategy of the *Heterorhabditis bacteriophora* mtDNA fragments sequenced in this study. The relative position of each PCR fragment is shown by a bold horizontal bar below the schematic linear gene map (genes are not drawn to scale and tRNA genes are not shown). Sequences of the PCR primers are listed in Supplementary Table S1.



Supplementary Fig. S2. ClustalX multiple sequence alignment of NCR3 from eight *Heterorhabditis* species: *H. bacteriophora*, *H. downesi*, *H. marelatus*, *H. megidis*, *H. zealandica*, *H. taysearae*, *H. mexicana* and *H. indica*. Conserved nucleotide positions are indicated by *.

Supplementary Table S1. Mitochondrial fragments and primer pairs used in this study for PCR amplification of the *Heterorhabditis bacteriophora* mitochondrial genome, as well as the studies from where the primer sequences were obtained.

Fragment name	Tm (°C)	Primer name	Primer sequence	Reference
Mt-1	*62	nad6F	5′-GAGTTATTTGAATTTTGATCCTATAAAGAGTAGG-3′	EST (Dolan, 2001)
	**68/4 min	atp6R	5′-CAAACTCAGCTAATTCATAAACTGCCTC-3′	EST (Dolan, 2001)
Mt-2	*62	atp6F	5′-GGTTAGAGAATTTTCTCGTCCTTTAGC-3′	EST (Dolan, 2001)
	**68/10 min	cox2R	5′-GTCCTCACGCTAAGACTGCC-3′	Present study
Mt-3	*62	cox2F	5′-GGCAGTCTTAGCGTGAGGAC-3′	Present study
	**68/10 min	nad5R	5′-GGAGTAGGCGCTCTTATAGC-3′	EST (Dolan, 2001)
Mt-4	*62	nad4F	5′-GGCTGCTTATTCTTCTGTTACCC-3′	Present study
	**68/10 min	nad5R	5′-GGAGTAGGCGCTCTTATAGC-3′	EST (Dolan, 2001)
Mt-5	*56	atp6F	5′-GGTTAGAGAATTTTCTCGTCCTTTAGC-3′	EST (Dolan, 2001)
	**68/10 min	MB39R (nad4)	5′-CAAAGAATAATAAAAAGATACCAA-3′	Liu *et al*., 1999
Mt-6	*56	nad5F	5′-GGTTTAAATTCTAAAGGGTTTAC-3′	EST (Dolan, 2001)
	**60/3 min	nad6R	5′-CATCTCAGAAATAACACCAG-3′	Present study

* Annealing temperature.
** Extension temperature/time.

Supplementary Table S2. Mitochondrial genome size and AT-rich region (length, location, and percentage A + T content) among some representative groups of chromadorean and enoplean nematodes.

Nematode species	GenBank accession number	Genome size (bp)	AT length (bp)	AT location (between)	A + T %	Reference
Class Chromadorea						
Order Rhabditida						
Heterorhabditis bacteriophora	NC_008534	18 128	2154	trnA-trnP	83.4	Present study
Heterorhabditis indica	NC_040293	18 128	2154	trnA-trnP	83.4	Mandadi *et al*. (2019), sequence from GenBank NCBI
Caenorhabditis elegans	NC_001328	13 794	466	trnA-trnP	93.1	Okimoto *et al*. (1992)
Ancylostoma duodenale	NC_003415	13 721	268	trnA-trnP	90.1	Hu *et al*. (2002)
Necator americanus	NC_003416	13 604	173	trnA-trnP	83.2	
Strongyloides stercoralis	NC_005143	13 758	469	trnD-nad1	85.0	Hu *et al*. (2003a)
Cooperia oncophora	NC_004806	13 636	304	trnA-trnP	85.5	Van der Veer & de Vries (2004)
Steinernema carpocapsae	NC_005941	13 925	554	trnW-trnE	79.2	Montiel *et al*. (2006)
Haemonchus contortus	NC_010383	14 055	465	trnA-trnP	89.4	Jex *et al*. (2008)
Steinernema glaseri	AP017464	13 851	1525	trnS2-trnN	84.4	Kikuchi *et al*. (2016)
Steinernema kushidai	AP017465	15 182	643	trnS2-trnT	79.7	
Steinernema litorale	AP017466	21 403	6260	trnN-trnT	72.3	
Ostertagia trifurcata	MK227249	14 151	113	trnR-trnV	80.1	Ahmad *et al*. (2019)
Order Ascaridida						
Ascaris suum	NC_001327	14 284	886	trnS2-trnN	84.7	Okimoto *et al*. (1992)
Anisaki simplex	NC_007934	13 916	515	nad4-cox1	87.2	Kim *et al*. (2006)
Ascaridia galli	NC_021642	13 977	610	trnC-trnN	80.0	Liu *et al*. (2013)
Ascaridia columbae	NC_021643	13 862	563	trnC-trnN	77.7	
Order Spirurida						
Onchocerca volvulus	NC_001861	13 747	312	cox3-trnA	85.3	Keddie *et al*. (1998)
Dirofilaria immitis	NC_005305	13 814	362	cox3-trnA	85.9	Hu *et al*. (2003b)
Brugia malayi	NC_004298	13 657	283	cox3-trnA	85.1	Ghedin *et al*. (2007)
Spirocerca lupi	KC305876	13 780	400	cox3-trnA	88.5	Liu *et al*. (2013b)
Camallanus cotti	NC_036308	17 901	237	trnH-trnR	77.2	Zou *et al*. (2017)

Supplementary Table S2. (Continued.)

Nematode species	GenBank accession number	Genome size (bp)	AT length (bp)	AT location (between)	A + T %	Reference
Order Oxyurida						
Enterobius vermicularis	NC_011300	14 010	675	trnS2-trnI	79.9	Kang *et al*. (2009)
Wellcomia siamensis	NC_016129	14 128	511	trnS2-trnN	92.3	Park *et al*. (2011)
Passalurus ambiguous	NC_028345	14 023	613	trnS2-trnN	91.7	Liu *et al*. (2016)
Order Tylenchida						
Pratylenchus vulnus	NC_020434	21 656	6847	cox1-trnH	72.9	Sultana *et al*. (2013)
Meloidogyne graminicola	NC_024275	20 030	5063	trnS2-trnM	84.1	Sun *et al*. (2014)
Meloidogyne arenaria	NC_026554	17 580	1023	trnD-trnM	80.5	Humphreys-Pereira & Elling (2015)
Meloidogyne javanica	NC_026556	18 291	1025	trnD-trnM	80.9	
Class Enoplea						
Order Dorylaimida						
Xiphinema americanum	NC_005928	12 626	95	nad3-nad4L	72.0	He *et al*. (2005)
Xiphinema pachtaicum	NC_033869	12 489	140	nad4L-nad3	75.0	Palomares-Rius *et al*. (2017)
Xiphinema rivesi	NC_033869	12 624	95	nad3-nad4L	73.0	
Longidorus vineacola	NC_033867	13 519	92	nad4L-nad3	70.0	
Order Mermithida						
Romanomermis culicivorex	NC_008640	26 194	7376	nad4-trnS2	77.6	Powers *et al*. (1993)
Thaumamermis cosgrovei	NC_008046	19-34 kb	401	nad1-cytb	73.0	Tang & Hyman (2007)
Hexamermis agrotis	NC_008828	24 606	3296	trnY-atp6	78.0	Tang *et al*. (2007), sequence from GenBank NCBI
Order Trichinellida						
Trichinella spiralis	NC_002681	16 706	1232	nad1-nad2	77.7	Lavrov & Brown (2001)
Trichuris discolor	NC_018596	13 904	126	nad1-trnK	74.6	Liu *et al*. (2012)
Trichinella nelsoni	NC_025753	15 278	722	nad1-trnK	74.7	Mohandas *et al*. (2014)

The sequences for *Heterorhabditis indica* (GenBank accession number NC_040293) and for *Hexamermis agrotis* (GenBank accession number NC_008828) were retrieved directly from NCBI.