

1-1-2011

## Poly(T) variation in heteroderid nematode mitochondrial genomes is predominantly an artefact of amplification

Angelique H. Riepsamen  
*University of Wollongong, angeliqu@uow.edu.au*

Tracey A. Gibson  
*University of Wollongong, tgibson@uow.edu.au*

Janet Rowe  
*Rothamsted Research, Harpenden, UK*

David J. Chitwood  
*United States Department of Agriculture*

Sergei Subbotin  
*California Department of Food and Agriculture*

*See next page for additional authors*

Follow this and additional works at: <https://ro.uow.edu.au/scipapers>



Part of the [Life Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Riepsamen, Angelique H.; Gibson, Tracey A.; Rowe, Janet; Chitwood, David J.; Subbotin, Sergei; and Downton, Mark P.: Poly(T) variation in heteroderid nematode mitochondrial genomes is predominantly an artefact of amplification 2011.  
<https://ro.uow.edu.au/scipapers/4077>

---

# Poly(T) variation in heteroderid nematode mitochondrial genomes is predominantly an artefact of amplification

## Abstract

We assessed the rate of in vitro polymerase errors at polythymidine [poly(T)] tracts in the mitochondrial DNA (mtDNA) of a heteroderid nematode (*Heterodera cajani*). The mtDNA of these nematodes contain unusually high numbers of poly(T) tracts, and have previously been suggested to contain biological poly(T) length variation. However, using a cloned molecule, we observed that poly(T) variation was generated in vitro at regions containing more than six consecutive Ts. This artefactual error rate was estimated at  $7.3 \times 10^{-5}$  indels/poly(T) tract >6 Ts/cycle. This rate was then compared to the rate of poly(T) variation detected after the amplification of a biological sample, in order to estimate the 'biological + artefactual' rate of poly(T) variation. There was no significant difference between the artefactual and the artefactual + biological rates, suggesting that the majority of poly(T) variation in the biological sample was artefactual. We then examined the generation of poly(T) variation in a range of templates with tracts up to 16 Ts long, utilizing a range of Heteroderidae species. We observed that T deletions occurred five times more frequently than insertions, and a trend towards increasing error rates with increasing poly(T) tract length. These findings have significant implications for studies involving genomes with many homopolymer tracts.

## Keywords

Polythymidine, Mitochondrial DNA, Heteroderidae, Polymerase slippage, Amplification artefacts, Homopolymer, Nematode, CMMB

## Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

## Publication Details

Riepsamen, A. H., Gibson, T., Rowe, J., Chitwood, D. J., Subbotin, S. A. & Dowton, M. (2011). Poly(T) variation in heteroderid nematode mitochondrial genomes is predominantly an artefact of amplification. *Journal of Molecular Evolution*, 72 (2), 182-192.

## Authors

Angelique H. Riepsamen, Tracey A. Gibson, Janet Rowe, David J. Chitwood, Sergei Subbotin, and Mark P. Dowton

# Poly(T) Variation in Heteroderid Nematode Mitochondrial Genomes is Predominantly an Artefact of Amplification

Angelique H. Riepsamen<sup>1</sup>, Tracey Gibson<sup>1</sup>, Janet Rowe, David J. Chitwood<sup>2</sup>, Sergei A. Subbotin<sup>3,4</sup>, and Mark Dowton<sup>1</sup>

<sup>1</sup>School of Biological Sciences, University of Wollongong, Wollongong, NSW, 2522, Australia; <sup>2</sup>United States Department of Agriculture, Agricultural Research Service, Nematology Laboratory, Beltsville, MD 20705, USA; <sup>3</sup>Plant Pest Diagnostics Center, California Department of Food and Agriculture, 3294 Meadowview Road, Sacramento, CA 95832-1448, USA; <sup>4</sup>Center of Parasitology of A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences, Leninskii Prospect 33, Moscow, 117071, Russia

## Abstract

We assessed the rate of in vitro polymerase errors at polythymidine [poly(T)] tracts in the mitochondrial DNA (mtDNA) of a heteroderid nematode (*Heterodera cajani*). The mtDNA of these nematodes contain unusually high numbers of poly(T) tracts, and have previously been suggested to contain biological poly(T) length variation. However, using a cloned molecule, we observed that poly(T) variation was generated in vitro at regions containing more than six consecutive Ts. This artefactual error rate was estimated at  $7.3 \times 10^{-5}$  indels/poly(T) tract >6 Ts/cycle. This rate was then compared to the rate of poly(T) variation detected after the amplification of a biological sample, in order to estimate the 'biological + artefactual' rate of poly(T) variation. There was no significant difference between the artefactual and the artefactual + biological rates, suggesting that the majority of poly(T) variation in the biological sample was artefactual. We then examined the generation of poly(T) variation in a range of templates with tracts up to 16 Ts long, utilizing a range of Heteroderidae species. We observed that T deletions occurred five times more frequently than insertions, and a trend towards increasing error rates with increasing poly(T) tract length. These findings have significant implications for studies involving genomes with many homopolymer tracts.

**Keywords** Polythymidine Mitochondrial DNA Heteroderidae Polymerase slippage Amplification artefacts Homopolymer Nematode

## Introduction

The mitochondrial genomes of bilaterian animals are typically single, ~16 kb circular molecules. They are inherited maternally and encode a highly conserved set of 37 mitochondrial genes, including 22 tRNAs, 2 rRNAs and 13 protein-coding genes. These mtDNAs are efficient molecules, generally having only a short non-coding region, no introns, and minimal to no intergenic sequences. Many genes overlap one another by several nucleotides (nts), such that post-transcriptional polyadenylation of the mRNA completes the transcript (Lavrov 2007). This succinctness has led to suggestions that larger mtDNA are selected against during evolution (Rand and Harrison 1989). These selective pressures may also contribute to the low frequency of inverted/tandem repeats or non-functional gene copies (pseudogenes) (Moritz et al. 1987).

However, complete sequencing of bilaterian animal mitochondrial genomes has revealed unusually large (>24 kb) mtDNA sporadically in a range of animals, primarily due to tandem repeats and lengthy non-coding regions. This includes in a fish (Inoue et al. 2010), a reptile (Fujita et al. 2007), molluscs (La Roche et al. 1990), nematodes (Azevedo and Hyman 1993; and GenBank accession no. EF368011), and an arthropod (Jeyaprakash and Hoy 2007). Pseudogenes of tRNAs, rRNAs and/or protein-coding genes have been characterised in the mtDNA of a range of organisms, including insects (Beckenbach et al. 2005; Dowton et al. 2003; Lessinger et al. 2004), amphibians (San Mauro et al. 2006), reptiles (Kumazawa et al. 1998; Russell and Beckenbach 2008; Zevering et al. 1991), fish (Mabuchi et al. 2004), birds (Mindell et al. 1998), a sponge (Rosengarten et al. 2008), an oyster (Milbury and Gaffney 2005), and in nematodes (Gibson et al. 2007a; Gibson et al. 2007b; Howe and Denver 2008). Long homopolymer

stretches have been reported in the coding regions of some animal mtDNA. Homopolymers are potentially detrimental as they are prone to transcriptional slippage, resulting in variation in the tract length relative to the template molecule (Wagner et al. 1990). Observations of decreased homopolymer tract lengths and frequencies in coding regions compared to non-coding regions have led to hypotheses of evolutionary selection against long (>~8 nts) homopolymer tracts in coding regions (Tamas et al. 2008). Despite this, long homopolymers are reported in the mtDNA of several invertebrate species, including a brachiopod, *Terebratalia transversa* (Helfenbein et al. 2001), and several nematodes, including *Dirofilaria immitis* (Hu et al. 2003b), *Haemonchus contortus* (Jex et al. 2008) *R. similis* (Jacob et al. 2009) and *Strongyloides stercoralis* (Hu et al. 2003a).

Recently we reported homopolymers in protein-coding mitochondrial genes of two potato cyst nematodes (*Globodera rostochiensis* and *Globodera pallida*) (Gibson et al. 2007a; Gibson et al. 2007b; Riepsamen et al. 2008). The mtDNA of these species are atypical, composed of several small circular mtDNA molecules (~6.4-9.5 kb in *G. pallida*), each with a large non-coding region and only some of the 37 mitochondrial genes in typical animal mtDNA (Armstrong et al. 2000). The gene organisation is an overlapping mosaic of large multigenic fragments present on several circles. These multigenic fragments have up to 98% identity on different circles for both *G. pallida* and *G. rostochiensis*. Several of these gene copies produced truncated proteins when conceptually translated, and hence were considered pseudogenes. Though pseudogenes are rare in animal mtDNA, duplicated gene copies are predicted to have reduced selective pressures and hence accumulate mutations. However, we observed the nucleotide sequence of these gene copies was neither divergent nor truncated, but had

indel mutations (insertions/deletions of 1-2 nucleotides) which disrupted the reading frame. Analysis of multiple clones revealed that these indels were consistently thymidine indels within poly(T) tracts on the coding strand (or adenine insertions/deletions within polyadenine [poly(A)] tracts on the non-coding strand). Termed 'poly(T) variation', these indels were not considered artefactual for several reasons; (i) both DNA strands were sequenced for all experiments; (ii) electropherograms showed no ambiguous base-calling surrounding the homopolymers; (iii) the DNA polymerase used for *G. rostochiensis* had proofreaders (both 5'→3' and 3'→5'); (iv) there was no evidence of polymerase errors, with substitution mutations observed between gene copies on different circles identical between *G. pallida* and *G. rostochiensis*; (v) the poly(T) variation was unrelated to experimental conditions, with reproduced amplifications on different populations having markedly different poly(T) variation, suggesting that poly(T) variation is population-specific; and (vi) poly(T) variation was characterised in amplicons from genomic DNA and mRNA (Riepsamen et al. 2008), and in EST databases generated independently ([www.nematode.net](http://www.nematode.net)).

Despite these observations, the possibility of poly(T) variation in *Globodera* mtDNA being an artefact of amplification could not be discounted, as for each instance an amplification preceded sequencing. Thus the poly(T) variation observed was more precisely a measure of biological variation plus any artefactual variation introduced (biological + artefactual variation). It is extremely difficult to measure biological poly(T) variation in nematodes without an amplification step. Multiple individuals cannot be used to increase the amount of biological material, as this may introduce inter-individual poly(T) variation. However, it is relatively simple to measure the artefactual variation (i.e. polymerase error) that a poly(T) tract generates *in vitro*. By

isolating a single DNA molecule from a region exhibiting poly(T) tracts, the amount of artefactual poly(T) variation can be measured by performing a second round of amplification, and sequencing multiple clones. If no polymerase errors occur, all copies amplified during this second round will be identical. If polymerase errors do occur and poly(T) variation is introduced, the proportion of clones with different poly(T) tract lengths will be a measure of the extent of artefactual poly(T) variation. As described above, the biological + artefactual variation can also be determined; i.e. the poly(T) variation observed after amplification of a biological sample (and the sequencing of multiple clones). By deducting the artefactual poly(T) variation from the biological + artefactual variation, the biologically-derived poly(T) variation can be estimated. The use of a single DNA molecule as the template for the analysis of artefactual variation avoids the possible introduction of biological variation, for example, due to mtDNA heteroplasmy. Further, by measuring the extent of both artefactual and artefactual + biological poly(T) variation after a single round of cloning, any variation introduced during cloning would be inherent in both poly(T) variation estimates. Thus, any differences between these estimates will primarily be due to amplification error.

In the present study, the extent of artefactual, and biological + artefactual poly(T) variation was determined, using a mtDNA region from a heteroderid nematode (the family to which *G. pallida* and *G. rostochiensis* belong). We investigated the generation of poly(T) variation using a long DNA polymerase, as ultimately we intended to use this to amplify long fragments of nematode mitochondrial genomes for sequencing. This revealed that the level of poly(T) variation characterised for the biological sample after an amplification step was not significantly greater than that introduced as an artefact of amplification. This suggests that poly(T) variation observed in the heteroderid mtDNA



is predominantly an artefact of amplification. The poly(T) variation generated during amplification of a broad range of heteroderid nematodes, as well as in two non-cyst-forming plant parasitic nematodes, was also studied to assess the potential each molecule has for poly(T) variation. These data were used to determine features that make these molecules susceptible to poly(T) variation. These results have significant implications for the sequencing of mitochondrial genomes containing extensive homopolymer regions. Unless strategies are applied to minimise or account for potential poly(T) variation, incorrect homopolymer tract lengths may be inferred, and genes can be incorrectly annotated as pseudogenes.

## Materials and Methods

### Genomic DNA Extraction and Molecular Identification

Nematode specimens were collected from Australian and international sources (see Table 5.1), and preserved in 95% ethanol on arrival. Single cysts of the *Heterodera* and *Betulodera* specimens were washed with distilled water and surface sterilised in 0.3% hypochlorous bleach as described (Nour et al. 2003). Sterilised cysts were washed with distilled water to remove any bleach prior to DNA extraction.

Table 1. Biological material used and origin of the specimens. All nematodes are of the superfamily Hoplolaimoidea.

Nematode	Source
Heteroderidae	
<i>Heterodera cajani</i>	Janet Rowe, UK

<i>Heterodera cardiolata</i>	Janet Rowe, UK
<i>Heterodera fici</i>	Janet Rowe, UK
<i>Heterodera glycines</i>	David Chitwood, USA
<i>Heterodera humuli</i>	Janet Rowe, UK
<i>Heterodera pratensis</i>	Janet Rowe, UK
<i>Heterodera zea</i>	David Chitwood, USA
<i>Betulodera betulae</i>	Sergei Subbotin, USA
Hoplolaimidae	
<i>Helicotylenchus dihystra</i>	Greg Walker, AUS
Meloidogynidae	
<i>Meloidogyne javanica</i>	Ian Riley, AUS

---

DNA extractions utilised (i) the egg mass of *M. javanica* within the gall formed on the infected root, (ii) an adult of *Helicotylenchus dihystra*, and (iii) sterilised cysts of the Heteroderidae nematodes. Cysts are formed from the posterior of a deceased gravid female. As the mtDNA of the eggs inside are maternally inherited (Tsang and Lemire 2002), this is the largest available mtDNA source representing a single individual. Total genomic DNA of all specimens were extracted using the salting out method described by Sunnucks and Hales (Sunnucks and Hales 1996), but with the proteinase K step extended to 1 h at 65°C, followed by 10 min at 95°C to inactivate the enzyme. Extracts were resuspended in TE buffer (1 mM Tris-HCl, 0.1 mM EDTA [pH 8]) and stored at 4°C until use. Species identifications were molecularly verified by amplification and sequencing across the ITS regions 1 and 2 using the Heteroderidae diagnostic primers TW81 (5'-GTTTCCGTAGGTGAACCTGC-3') and AB28 (5'-ATATGCTTAAGTTCAGCGGGT-3') as described by Subbotin *et al.* (2001). Amplification reactions used the BIO-X-ACT™ Long DNA Polymerase Kit (Bioline,

Australia), and cycling conditions consisted of 94°C for 2 min, 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min, with a final elongation of 72°C for 5 min. Sequencing reactions were carried out using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Australia), and an ABI 3130xl Genetic Analyser (Applied Biosystems, Australia). Sequences were analysed using ChromasPro (Technelysium Ltd., Tewantin, Australia), BioEdit (Hall 1999) and ClustalW (Thompson et al. 1994) software. The RNAfold software (<http://rna.tbi.univie.ac.at>) was used to predict secondary structures in the DNA sequences (Hofacker 2003).

### Amplification and Cloning of a Mitochondrial Fragment

A fragment of the *COII* gene containing poly(T) tracts was amplified in *Heterodera cajani* using the BIO-X-ACT™ Long DNA Polymerase Kit, and primers Het-COII-F (5'-GTGGGTCATCARTGRTAYTG-3') and Het-COII-R (5'-GAATGATTAGCTCCRCARATYTC-3'). PCR cycling conditions consisted of 94°C for 2 min, 35 cycles of 94°C for 10 s, 45-65°C for 30 s and 68°C for 1 min/kb, followed by 68°C for 5 min. The PCR product was cloned directly into the pGEM®-T Easy Vector System (Promega, Australia). Two positive clones were purified using the Wizard® Plus SV Minipreps DNA Purification System (Promega). Both DNA strands were sequenced using the sequencing and analysis methods described above, using M13 vector primers.

### Characterisation of Artefactual (*In Vitro*) Variation at Poly(T) Tracts

The COII fragment of the purified plasmid DNA (representing a single DNA molecule) was subcloned to facilitate analysis of the poly(T) variation generated *in vitro*, as per the

approach of Clarke *et al.* (2001). Re-amplification and cloning used the same primers, enzyme, cycling conditions and vector system as described above. Two approaches were used to enable sequencing of these clones. In the first approach, purified products of colony-PCRs were used for sequencing. Colony-PCRs used the BIO-X-ACT™ Long DNA Polymerase Kit as described above, with products purified using ExoSAP-IT® (USB Corporation, Australia). In the second approach, plasmid DNA was purified for sequencing using the Wizard® Plus SV Minipreps DNA Purification System. For both approaches sequencing was performed on both DNA strands using M13 primers as described above.

## Characterisation of Total (Biological + Artefactual) Variation at Poly(T) Tracts

The same COII region was analysed in a broad range of species to assess the features that may contribute to poly(T) variation. For *H. cajani*, *Heterodera cardiolata*, *Heterodera glycines* and *Heterodera zaeae*, species-specific primers P116F-1R (5'-CCACGGGTATAAAGGAATGA-3') and H.Gly-COII<sub>F</sub>\_inFOR (5'-GGTAGTTGGTCATCAATGATATTG-3') were used. To optimise amplification from divergent species, degenerate primers Het-COII-F and Het-COII-R (described above) were used for all species listed in Table 5.1. Both primer pairs encompass the same COII region. Amplification, cloning, and sequencing of multiple clones from the purified colony-PCR product were performed as described above.

An internal fragment of the *COI* gene was also amplified in *H. cardiolata* & *H. glycines* using the primers H.Gly-COII<sub>inFOR</sub>-4F (5'-TGCTTTAATTGGTGCTTTTGG-3') and

H.Gly-COI-FOR (5'-TATTCCAACAACAAATATATGATGAGC-3'). For *H. glycines* the entire *COI* gene was further amplified, both as two overlapping fragments and as one larger fragment encompassing the entire gene. The primer pairs for these amplifications were H.Gly-COIFOR and H.Gly-COIIinFOR (5'-GGTAGTTGGTCATCAATGATATTG-3') for the 5' fragment, H.Gly-COI-REV (5'-CTGTTTTGGCAGGAGCTATTAC-3) and H.Gly-ND4inREV1 (5'-GAATTAATTTTTCAATTTGGACTCC-3') for the 3' fragment, and H.Gly-COIIinFOR-2F (5'-GCACCTTTAGCGTGATTGGT-3') and H.Gly-ND4inREV1-2F (5'-TGAAAGCCAACTAAAGCCTCA-3') for the entire gene. The methods for amplification, cloning, sequencing from the colony-PCR product and analysing multiple clones were as described above for the COII fragment.

## Results

### Artefactual Poly(T) Variation

A region of the *COII* gene in *H. cajani* was amplified and the PCR product sequenced directly. The electropherogram showed evidence of heterogeneity (see Appendix 5.1). This heterogeneity was evident as peaks of lesser intensity than the main peak, and was always seen immediately downstream (3') of a poly(T) region. In each case, the minor peak corresponded to the next-most nucleotide, suggesting that the secondary peaks were the product of slippage errors. To determine the extent to which poly(T) variation was generated *in vitro*, individual (i.e. cloned) molecules of a fragment containing poly(T) tracts were characterised. A 274 bp region of the *COII* gene from *H. cajani* was sequenced in two clones (denoted *a1* and *a2*). The electropherograms of these

sequences were clean, with no evidence of heterogeneity (for an example, see Appendix 5.1), suggesting negligible variation was introduced by cloning. The COII sequences of *a1* and *a2* were identical and translated a conserved polypeptide. To represent single DNA molecules, plasmid DNA of *a1* and *a2* was purified. This DNA was subject to a second round of amplification and cloning. Ten clones were sequenced from each to assess the level of variation introduced during this second round of amplification, cloning and sequencing (i.e. artefactual variation). However, in order to sequence these clones the template had to be amplified, either by culture of the bacterial clones ('mini-prep'), or by colony-PCR. Sequences obtained from individual clones using either approach were identical (data not shown). Examination of the electropherograms of each clone sequenced using either approach showed no ambiguous peaks as to suggest that either the colony PCR or the bacterial propagation were individually contributing to the sequence reported (for an example, see Appendix 5.1). Thus, although some amplification errors may have been introduced during either the colony PCR or the bacterial propagation, this did not influence the result, with each reported sequence representing a PCR amplicon. Thus, the more rapid approach of sequencing from the colony-PCR product was used for subsequent experiments.

Comparison of the 20 cloned sequences from amplification of *a1* and *a2* revealed that several poly(T) indels and substitution mutations were introduced *in vitro* (Table 5.2). Across the poly(T) tracts in this COII template, 4 clones had a single T deletion (-1T); three at a tract of 10 Ts (Fig. 5.1), and one at a tract of 9 Ts. This resulted in a poly(T) variation rate of  $1.90 \times 10^{-3}$  indels/poly(T) tract >6 Ts/cycle. In this study, indels at poly(T) tracts were only observed at tracts longer than 6 Ts. Thus, this rate will hereafter be referred to as indels/poly(T)/cycle. Substitution mutations were present in

10 clones, and included 13 transition mutations and one transversion, giving a substitution mutation rate of  $7.30 \times 10^{-5}$  mutations/bp/cycle (Table 5.2).

Table 2. Poly(T) variation and substitution mutation rates within a ~274 bp COII fragment amplified and cloned from both individual molecules isolated *in vitro* (i.e. artefactual variation) and from biological samples from a range of nematode species (i.e. biological + artefactual variation).

COII template DNA	Number of clones analysed	Number of poly(T) tracts >6 Ts <sup>a</sup>	Poly(T) variation rate <sup>b</sup> (x10 <sup>-3</sup> )	Substitution mutation rate <sup>c</sup> (x10 <sup>-5</sup> )
Artefactual	20	3	1.90	7.30
Artefactual + Biological				
Heteroderidae				
<i>Heterodera cajani</i>	35	3	2.18	2.98
<i>Heterodera cardiolata</i>	22	2	0.65	0.95
<i>Heterodera fici</i>	10	5	1.71	2.09
<i>Heterodera glycines</i>	15	2	0.95	0.70
<i>Heterodera humuli</i>	10	5	2.29	2.09
<i>Heterodera pratensis</i>	9	5	2.54	0.00
<i>Heterodera zaeae</i>	14	5	2.45	0.74
<i>Betulodera betulae</i>	10	3	0.95	1.04
Hoplolaimidae				
<i>Helicotylenchus dihystrera</i>	10	1	0.00	1.04
Meloidogynidae				
<i>Meloidogyne javanica</i>	10	1	0.00	1.04

<sup>a</sup>Per template molecule. Only tracts with >6 Ts demonstrated poly(T) variation.; <sup>b</sup>Poly(T) variation rates (indels/poly(T)/cycle) were calculated as  $im/(t \times c)$ , where  $im$  is the number of indel mutations,  $t$  is the total number of poly(T) sites longer than 6 Ts sequenced, and  $c$  is the number of amplification cycles;

<sup>c</sup>Substitution mutation rates (mutations/bp/cycle) were calculated as  $sm/(bp \times c)$ , where  $sm$  is the number of substitution mutations,  $bp$  is the total number of base pairs sequenced, and  $c$  is the number of amplification cycles.



```

a1 TL  I I D F F F  E A T G V F Y
      230      240      250      260
      ./.../.../.../.../.../.../.../...
a1    ATTATTGATTTTTTTT-GAGGCTACCGGAGTTTTTAT
a1.1  .....
a1.2  .....
a1.3  .....
a1.4  .....
a1.5  .....
a1.6  .....
a1.7  .....
a1.8  .....
a1.9  .....
a1.10 .....
a2    .....
a2.1  .....
a2.2  .....
a2.3  .....
a2.4  .....
a2.5  .....
a2.6  .....T.....
a2.7  .....
a2.8  .....
a2.9  .....
a2.10 .....A.....
bD1   .....T.....
bD2   .....
bD3   .....T.....
bD4   .....
bD5   .....
bD6   .....
bD7   .....
bD8   .....
bD9   .....
bD10  .....
bD11  .....
bD12  .....
bD13  .....
bD14  .....
bD15  .....
bD16  .....
bD17  .....
bS1   .....T.....
bS2   .....
bS3   .....
bS4   .....
bS5   .....
bS6   .....
bS7   .....
bS8   .....
bS9   .....
bS10  .....
bS11  .....T.....G
bS12  .....
bS13  .....
bS14  .....
bS15  .....
bS16  .....

```

Figure 1 Section of an alignment of the cloned *COII* gene fragment amplified from *Heterodera cajani*. Clones *a1.1-a1.10* and *a2.1-a2.10* were each amplified from a single molecule, thus any sequence variation is an artefact of amplification. The remaining clones were amplified from genomic DNA and thus represent biological + artefactual variation, with clones *a1*, *a2* and *bD1-bD17* amplified using a degenerate primer set, and clones *bS1-bS16* amplified using a specific primer set. T<sub>L</sub> indicates the conceptual translation of the *a1* clone.

## Biological + Artefactual Poly(T) Variation

The extent of biological + artefactual poly(T) variation was then assessed in the same COII region of *H. cajani* (i.e. cloned COII fragments amplified from genomic DNA templates). Both specific and degenerate primer pairs were used to amplify this COII fragment, from which 16 and 17 clones were sequenced (denoted *bS1-bS16* and *bD1-bD17*, respectively in Fig. 5.1). Poly(T) length variation was present in 8 of the 33 clones; at a tract of 10 Ts four clones had -1T and two clones had a single T insertion (+1T) (Fig. 5.1), and at a tract of 9 Ts two clones had -1T. This gives a poly(T) variation rate of  $2.18 \times 10^{-3}$  indels/poly(T)/cycle (Table 5.2). Substitution mutations were present in 9 clones, and included 6 transition mutations and 4 transversions, giving a substitution rate of  $2.98 \times 10^{-5}$  mutations/bp/cycle (Table 5.2).

A 2-tailed chi-square test was performed to investigate whether the biological + artefactual poly(T) variation was significantly different to the artefactual poly(T) variation ([www.graphpad.com/quickcalcs/contingency1.cfm](http://www.graphpad.com/quickcalcs/contingency1.cfm)). There was no significant difference ( $p = 0.985$ ) between the rates of artefactual poly(T) variation and the artefactual + biological poly(T) variation in the clones analysed for this COII fragment from *H. cajani*. There was also no significant difference ( $p = 0.061$ ) between the artefactual substitution rate and the artefactual + biological substitution rate. This suggests that the majority of both the poly(T) variation and substitutions observed in the biological sample are artefacts of amplification. However, a more extensive investigation may be required to detect low levels of biological poly(T) variation at statistically significant levels.

## Template-Specific Poly(T) Variation in a *COII* Gene Fragment

To determine the sequence characteristics associated with the generation of poly(T) variation, this same COII region was cloned from a range of plant parasitic nematodes of the Hoplolaimoidea superfamily (Table 5.1). For all nematodes studied, the 274 bp (271 bp in *B. betulae*) COII fragment described for *H. cajani* was analysed. For each species 9-22 clones were characterised (Table 5.2). In each case, intraspecific sequence identity was 98-100%, suggesting that divergent numts (nuclear copies of mitochondrial DNA) were not being co-amplified and analysed. Interspecific sequence identity ranged from 63-85%, providing diverse template molecules for the characterisation of the generation of poly(T) variation. Despite 14 instances of poly(T) tracts 6 nucleotides long, variation was only evident at tracts >6 Ts. Only the Heteroderidae family displayed poly(T) variation, though template molecules of the Hoplolaimidae and Meloidogynidae families had only one poly(T) tract >6 Ts, compared with 2-5 for the Heteroderidae species (Table 5.2). Across the Heteroderidae, the rate of poly(T) variation ranged from  $0.65 \times 10^{-3}$  to  $2.54 \times 10^{-3}$  indels/poly(T)/cycle (Table 5.2). Variation consisted of -1T at 20 poly(T) tracts (ranging from 7-11 Ts), and a deletion of 2 Ts (-2Ts) at two tracts (of 9 and 10 Ts). It is difficult to predict whether the -2Ts is the result of two single T deletion events or one deletion event involving two Ts.

All species except *H. pratensis* displayed intraspecific nucleotide substitutions, with substitution rates ranging from  $0.70 \times 10^{-5}$  to  $2.98 \times 10^{-5}$  mutations/bp/cycle (Table 5.2).

## Template-Specific Variation in a *COI* Gene Fragment

A 625 bp central region of the *COI* gene in *H. cardiolata* and *H. glycines* was also amplified, sequenced and cloned. None of the clones displayed poly(T) variation, despite the template molecules having 3 and 2 poly(T) tracts of 7 Ts, respectively. Of the 10 clones sequenced for *H. cardiolata*, there was one transversion mutation, giving a substitution rate of  $0.64 \times 10^{-5}$  mutations/bp/cycle. Of the 11 clones sequenced for *H. glycines*, there were two substitution mutations (one transition and one transversion), giving a substitution rate of  $1.16 \times 10^{-5}$  mutations/bp/cycle.

## Template-Specific Poly(T) Variation Across the Entire *COI* Gene

The entire *COI* gene was amplified and cloned in *H. glycines*, both as two overlapping fragments, and as a larger fragment encompassing the entire *COI* gene. *COI* was sequenced in two clones and determined to be 1524 bp long. Start and stop codons were predicted as GTT and TAA, respectively. The two overlapping fragments characterised spanned 883 bp of the 5' end and 937 bp of the 3' end. Two clones of each overlapping fragment were also sequenced, thus each nucleotide of *COI* was sequenced in at least 4 clones, with the central portion sequenced in 6 clones.

There were 18 poly(T) tracts across the *COI* gene, ranging in length from 7 to 16 Ts. When all clones were compared, there were 13 instances of poly(T) variation across all tracts with 8 or more Ts. These included six tracts with -1T, three with +1T, two with -2Ts, and two with insertions of 2Ts (+2Ts) (based on the *COI* sequence translating the correct reading frame). This resulted in a poly(T) variation rate of  $5.02 \times 10^{-3}$

indels/poly(T)/cycle. This is 2-fold higher than any values reported for COII, though this may have been skewed by the relatively low number of clones analysed.

Across all clones, there were 5 substitution mutations (all transitions), giving a substitution rate of  $2.14 \times 10^{-5}$  mutations/bp/cycle.

### Features of Sites Prone to Poly(T) Length Variation

Analysis of the DNA sequences both up- and downstream of poly(T) tracts did not locate any sequence motifs or potential secondary structures associated with the poly(T) variation observed. If poly(T) variation was associated with a sequence motif, closely related species (with more similar DNA sequences) would be expected to have more similar poly(T) variation rates. However, there was no apparent correlation between poly(T) variation and phylogenetic proximity. For example, *H. cajani* and *H. glycines* are two of the more closely related species studied (both of the Schachtii clade of *Heterodera* nematodes) (Subbotin et al. 2001), yet have differing poly(T) variation rates ( $2.18 \times 10^{-3}$  and  $0.95 \times 10^{-3}$  indels/poly(T)/cycle, respectively; Table 5.2). Further, there was no correlation between clones with poly(T) variation and those with nucleotide substitutions.

The only characteristic common to all sites displaying poly(T) variation was the length of the poly(T) tract. Comparison of poly(T) tract lengths and poly(T) variation rates across all clones for all indel types (Fig. 5.2) identified no variation at any tract with 6 or fewer Ts, and a linear increase in variation rate with increased poly(T) tract length for tracts 7-16 Ts long. Observed individually, only the -1T indels appear to follow this

trend (Appendix 5.2). However, the distribution of other indels (+1T, -2Ts and +2Ts) at poly(T) tracts is more sporadic, and there are insufficient occurrences to establish a trend.

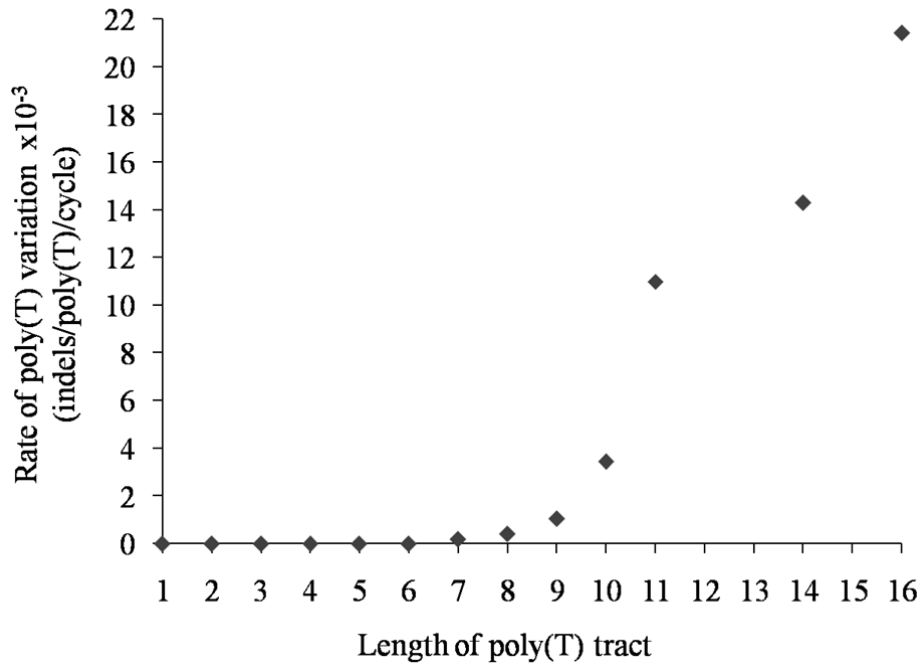


Figure 2 Relationship between the rate of poly(T) variation and poly(T) tract length of all indel types in COI and COII for all Hoplolaimoidea species studied.

Of the 630 poly(T) tracts that were longer than 6 Ts analysed across all COI and COII clones, the overall rate of poly(T) variation was  $1.95 \times 10^{-3}$  indels/poly(T)/cycle. Of these, 74% were single T deletions, 12% were single T insertions, 9% were deletions of 2 Ts, and 5% were insertions of 2 Ts (Appendix 5.2).

## Discussion

### Features Associated with Poly(T) Sites

Analysis of the generation of poly(T) variation *in vitro* revealed that the majority of poly(T) variation was artefactual, not biological. Further sequence analysis using multiple clones of *COI* and *COII* gene fragments in the mtDNA of Heteroderidae nematodes revealed poly(T) variation throughout many poly(T) tracts in these template molecules. This allowed the characteristics associated with poly(T) variation to be assessed. Single T deletions were the most frequent type of poly(T) variation, accounting for 74% of the total poly(T) variation observed. Overall, contractions were 5 times more frequent than expansions. Studies of homopolymers in nuclear DNA have found similar trends, with contractions 5-10 times more frequent than expansions (Kroutil et al. 1996; Kunkel 1986; Shinde et al. 2003; Tindall and Kunkel 1988). It was proposed that this may be due to contractions being more energetically favourable than expansions (Kroutil et al. 1996; Shinde et al. 2003). These studies predict a simple slippage model involving one base per indel event. Our observations support this model, with indels involving one nucleotide 6-fold more frequent than those involving two. Thus, the observed indels of -2Ts are likely the result of two subsequent single indel events. Further, poly(T) variation rates were found to increase with increasing tract lengths. Similar trends of increased mutation rate with increasing number of repeats have been reported for monomeric A/T microsatellites (Bacon et al. 2001; Lai and Sun 2003; Shinde et al. 2003) and for short poly(T) tracts up to 5 nts (Kroutil et al. 1998). If these errors arise by slippage of the polymerase or template molecule during replication, longer homopolymer runs may stabilise (and hence maintain) the mutation more efficiently (Kunkel 1990). The binding capacity of the polymerase active site is also

thought to be significant. Typically there are 7-9 nts in the active site during replication, however this can increase to up to 12 nts (Eom et al. 1996). Shinde *et al.* (2003) suggest that saturation of the active site with repetitive sequences may also contribute to an increased mutation rate with increasing number of repeats. Despite the present study demonstrating a considerably higher rate of poly(T) variation for tracts  $\geq 11$  Ts, there are too few instances to support a correlation between this and the nucleotide binding capacity of the polymerase enzyme.

### Consideration of Poly(T) Errors Introduced *In Vitro*

The majority of genomic DNA studies involve *in vitro* DNA amplification using a polymerase. These studies often require downstream cloning and expression experiments which rely on accurate DNA replication during amplification. However, the generation of amplification artefacts *in vitro* is often understated despite the potential implications of not accounting for artefacts if they are introduced.

#### *Polymerase error rates associated with homopolymer tracts.*

Studies of homopolymer polymerase errors have estimated error rates at  $1.5 \times 10^{-2}$  mutations/repeat/cycle in non-coding microsatellites (Shinde et al. 2003) and between  $0.42 \times 10^{-6}$  and  $90 \times 10^{-6}$  mutants/colony-forming unit/detectable site in a bacterial gene, *lacZ* (Kroutil et al. 1998). However, the differences in methodology between these studies and the present study (which estimated  $1.9 \times 10^{-3}$  indels/poly(T)/cycle) are significant and confound rate comparisons. The studies differed in the type of polymerase and template molecule used, and in the mutation detection and rate



estimation methods. For example, the present study estimated polymerase errors from a proofreading enzyme mix across poly(T) tracts up to 16 Ts long in isolated clones derived from biological templates. In comparison, Shinde *et al.* (2003) assessed the mutation rate of *Taq* polymerase (i.e. without a proofreader) across plasmids containing synthesised poly(A) inserts of 8-12 As. Poly(A) variation was detected via analysis of ‘stutter peaks’ associated with the primary peak during capillary electrophoresis of fluorescently-labelled PCR products. Mutation rates were estimated from this using a complex mathematical model that accounted for replication variables associated with (for example) increasing cycle number. Further, Kroutil *et al.* (1998) assessed ‘gap-filling’ *Taq* polymerase error rates at poly(T) tracts up to 5 Ts long. Using the bacterial *lacZ* gene, mutants were detected by blue/white colony screening and sequence analysis. Thus, comparisons with other studies using long-PCR polymerase mixes may be more appropriate (although these have not examined the mutation rate at homopolymers). Mutation rates of *Taq/Pfu* and *Klentaq/Pfu* polymerase mixes were reported as  $5.6 \times 10^{-6}$  and  $3.9 \times 10^{-6}$  mutations/bp/duplication, respectively (Cline *et al.* 1996). These rates are approximately three orders of magnitude less than the poly(T) error rate observed in the present study, and one order of magnitude less than the *in vitro* substitution rate in the present study ( $7.30 \times 10^{-5}$  mutations/bp/cycle). This further indicates the propensity towards polymerase errors at poly(T) tracts relative to other DNA sequences. Regardless, there remains a need to identify genomes susceptible to *in vitro* homopolymer length variation.

### *Genomes susceptible to polymerase indel errors*

Analysis of the mtDNA sequences in GenBank from other species (Table 5.3) demonstrates a high frequency of poly(T) tracts >6 Ts in many other parasitic nematodes, but not in free-living nematodes, such as *Caenorhabditis elegans*. The mitochondrial genomes of the parasitic nematodes are therefore at high risk of indel errors being introduced *in vitro*. Characterising the mtDNA of parasitic nematodes has previously been described as difficult, suspected to be a result of their high AT content (Hu et al. 2007). Poly(A/T) tracts contribute to this high AT content. For example, *R. similis* and *H. glycines* of the Chromadorea class of nematodes have the highest AT contents of any animal mtDNA sequenced to date, with 85.4% (Jacob et al. 2009) and 82.6 % AT content (Gibson et al. *submitted*), respectively (cf. 76.2% in *C. elegans*; GenBank accession no. NC\_001328). *R. similis* and *H. glycines* also have some of the highest incidence of poly(T) tracts, with 161 and 347 poly(T) tracts >6 Ts, respectively (cf. 28 in *C. elegans*; Table 5.3). With poly(T) tracts so prevalent in parasitic nematodes it is vital that the potential for poly(T) variation be recognised both during annotation, and during cloning and expression experiments. The annotation of *R. similis* mtDNA utilised sequence data from both direct sequencing of genomic DNA and from sequencing of plasmid DNA containing EST sequences derived from mRNA. Different polymerases were used for each. Poly(T) variation was identified and the biological poly(T) length predicted from comparison of corresponding genomic DNA and EST sequences. Thus, for analysing genomes with many homopolymers, methodologies need to be scrutinised to ensure they factor in the potential for homopolymer length variation.

Table 3. Incidence of poly(T) tracts longer than 6 nucleotides in the entire mtDNA of several animal species described as having many homopolymers, and relatives of these species also found to have many poly(T) tracts. The free-living nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* are included for comparison, as all other nematode representatives have a parasitic life stage.

Species	Number of poly(T) tracts >6 Ts in mtDNA	GenBank accession no.
Brachiopoda		
<i>Terebratalia transversa</i>	39	AF331161
Nematoda		
<i>Heterodera glycines</i>	347 <sup>a</sup>	HM640930
<i>Strongyloides stercoralis</i>	171	NC_005143
<i>Dirofilaria immitis</i>	167	NC_005305
<i>Radopholus similis</i>	161	FN313571
<i>Brugia malayi</i>	139	NC_004298
<i>Onchocerca volvulus</i>	123	NC_001861
<i>Steinernema carpocapsae</i>	119	NC_005941
<i>Mecistocirrus digitatus</i>	59	GQ888722
<i>Haemonchus contortus</i>	50	EU346694
<i>Caenorhabditis elegans</i>	28	NC_001328
<i>Caenorhabditis briggsae</i>	5	NC_009885

*Strategies for minimising polymerase errors and their potential impact.*

Strategies for minimising *in vitro* errors should be considered when amplifying from genomes suspected of having many homopolymers. For amplification across homopolymers, the choice of polymerase has been shown to have a significant impact on fidelity (Fazekas et al. 2010). DNA polymerases *Bst* and *Tfu* have demonstrated resilience to replication slippage, suggested to be a factor of the strand-displacement

activity of the enzyme (Viguera et al. 2001). Fusion polymerases, such as Phusion™ and Herculase® II Fusion, have also shown improved sequencing quality over other polymerases. The additional binding domain of fusion polymerases is thought to increase the contact between the polymerase and template molecules, and hence improves stability and fidelity during replication (Fazekas et al. 2010). Although including a polymerase with a proofreader can improve polymerase fidelity 7-fold (Bebenek and Kunkel 1990), Clarke *et al.* (2001) found no difference in *in vitro* homopolymer errors using *Taq* and *Pfu* DNA polymerases. Further studies are required to determine the optimal polymerase, or combination of polymerases, for accurately amplifying genomes with long homopolymer tracts. Increasing the amount of template DNA in the reaction and decreasing the number of replication cycles will allow similar quantities of product to amplify, while decreasing the number of amplifications, minimising the impact of length variation introduced early in the PCR. This strategy is further supported by observations that mutations occur more frequently during later PCR cycles (Eckert and Kunkel 1991). During sequence analysis, homopolymer tract lengths should be interpreted carefully. To ensure that all tract lengths are accurate, amplification and sequencing should not rely on a single approach.

### *Sequencing strategies for genomes with extensive homopolymer tracts.*

There is no error-free method currently available for sequencing across homopolymer tracts. Direct sequencing across poly(T) regions using amplification products can be difficult, particularly when longer tracts are concerned. As the length of the tract increases, the rate of indel errors increases, and thus the proportion of length variants that are sequenced. This is of particular concern when the error is introduced during an

early cycle of the PCR, as it will be a significant component of the final product. During sequencing, this causes mixed signals in the sequence data downstream of poly(T) regions. This becomes progressively worse as the number of poly(T) tracts in the sequence increases. A study by Madani *et al.* (2010) encountered this problem when sequencing the mtDNA *cytochrome b* gene of *G. pallida* and *G. rostochiensis*. They were forced to clone the PCR product and sequence from the plasmid DNA of a single colony to obtain readable sequence data. This ensured that only one DNA amplicon was being sequenced. However, poly(T) variation was still present in many of the cloned sequences, preventing conceptual translation into a functional protein. If poly(T) variation is introduced during early PCR cycles, they may still form a considerable portion of the amplified molecules, and it becomes a matter of chance as to whether the plasmid sequenced contains DNA representative of the biological poly(T) tract length. Further, substitution errors introduced *in vitro* may also be cloned and sequenced. Thus it is necessary to sequence several clones to ensure the accuracy of the sequence data. Next generation sequencing technologies, such as the 454 sequencing platform, also have difficulties accurately sequencing tracts >6 Ts. These technologies are based on the detection and analysis of light emissions that are proportional to the number of identical nucleotides incorporated adjacently. However, for long homopolymers the detector becomes saturated with light and the base-calling software can no longer accurately interpret the amount of light emitted. This is predicted to occur when homopolymer tracts of 7 or more nucleotides are sequenced (Mardis 2008). Contrary to this, a study by Jex *et al.* (2010) evaluated the use of 454 technology to sequence the mtDNA of a chromadorean nematode, *Haemonchus contortus*, which has an AT content of 78.1% and poly(T) tracts up to 9 Ts. Although they observed many errors associated with homopolymers, the final 14,055 bp mtDNA consensus sequence was >99.7%

accurate. Thus, the 40-fold depth of coverage achieved with the high throughput analysis of 454 generally outweighed the frequency of base-calling errors. However, this sequencing approach was still preceded by PCR amplification from genomic DNA. Thus, sequencing methods that avoid amplification altogether would be superior. However, the longest poly(T) tract of *H. contortus* mtDNA is only 9 nt. Thus, although this singular approach appears sufficient, it may not be suitable for the mtDNA of genomes with longer homopolymer tracts, such as that of the Tylenchida order of nematodes to which *R. similis* and *H. glycines* belong. (See Appendix 5.3 for the frequencies of different tract lengths for the nematode species listed in Table 3).

These observations have significant implications for amplifying, sequencing, annotating and protein expression studies. The potential for the generation of poly(T) variation should be a consideration for all studies involving genes with many homopolymers longer than 6 nucleotides.

## References

- Armstrong MR, Blok VC, Phillips MS (2000) A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics* 154:181–192
- Azevedo JLB, Hyman BC (1993) Molecular characterization of lengthy mitochondrial DNA duplications from the parasitic nematode *Romanomermis culicivorax*. *Genetics* 133:933–942
- Bacon AL, Dunlop MG, Farrington SM (2001) Hypermutability at a poly(A/T) tract in the human germline. *Nucl Acids Res* 29: 4405–4413
- Bebenek K, Kunkel TA (1990) Frameshift errors initiated by nucleotide misincorporation. *Proc Natl Acad Sci USA* 87: 4946–4950
- Beckenbach AT, Robson SKA, Crozier RH (2005) Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome b gene in ants of the genus *Polyrhachis*. *J Mol Evol* 60: 141–152
- Clarke LA, Rebelo CS, Goncalves J, Boavida MG, Jordan P (2001) PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. *Mol Pathol* 54:351–353
- Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucl Acids Res* 24:3546–3551
- Dowton M, Castro LR, Campbell SL, Bargon SD, Austin AD (2003) Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. *J Mol Evol* 56:517–526
- Eckert KA, Kunkel TA (1991) DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl* 1:17–24
- Eom SH, Wang J, Steitz TA (1996) Structure of Taq polymerase with DNA at the polymerase active site. *Nature* 382:278–281
- Fazekas AJ, Steeves R, Newmaster SG (2010) Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques* 48:277–285
- Fujita MK, Boore JL, Moritz C (2007) Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). *Mol Biol Evol* 24:2775–2786
- Gibson T, Blok VC, Dowton M (2007a) Sequence and characterization of six mitochondrial subgenomes from *Globodera rostochiensis*: multipartite structure is conserved among close nematode relatives. *J Mol. Evol.* 65:308–315
- Gibson T, Blok VC, Phillips MS, Hong G, Kumarasinghe D, Riley IT, Dowton M (2007b) The mitochondrial subgenomes of the nematode *Globodera pallida* are mosaics: evidence of recombination in an animal mitochondrial genome. *J Mol Evol* 64: 463–471

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95–98

Helfenbein KG, Brown WM, Boore JL (2001) The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. *Mol Biol Evol* 18:1734–1744

Hofacker IL (2003) Vienna RNA secondary structure server. *Nucl Acids Res* 31:3429–3431

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

Hu M, Chilton NB, Gasser RB (2003a) The mitochondrial genome of *Strongyloides stercoralis* (Nematoda)—idiosyncratic gene order and evolutionary implications. *Int J Parasitol* 33:1393–1408

Hu M, Gasser RB, Abs El-Osta YG, Chilton NB (2003b) Structure and organization of the mitochondrial genome of the canine heartworm, *Dirofilaria immitis*. *Parasitology* 127:37–51

Hu M, Jex AR, Campbell BE, Gasser RB (2007) Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. *Nat Protoc* 2:2339–2344

Inoue JG, Miya M, Lam K, Tay B-H, Danks JA, Bell J, Walker TI, Venkatesh B (2010) Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol Biol Evol* 27:2576–2586

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

Jex AR, Hu M, Littlewood DT, Waeschenbach A, Gasser RB (2008) Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). *BMC Genomics* 9:11

Jex AR, Littlewood DT, Gasser RB (2010) Toward next-generation sequencing of mitochondrial genomes-focus on parasitic worms of animals and biotechnological implications. *Biotechnol Adv* 28: 151–159

Jeyaprakash A, Hoy MA (2007) The mitochondrial genome of the predatory mite *Metaseiulus occidentalis* (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. *Gene* 391:264–274

Kroutil LC, Register K, Bebenek K, Kunkel TA (1996) Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry* 35:1046–1053

Kroutil LC, Frey MW, Kaboord BF, Kunkel TA, Benkovic SJ (1998) Effect of accessory proteins on T4 DNA polymerase replication fidelity. *J Mol Biol* 278:135–146

Kumazawa Y, Ota H, Nishida M, Ozawa T (1998) The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150:313–329



- Kunkel TA (1986) Frameshift mutagenesis by eucaryotic DNA polymerases in vitro. *J Biol Chem* 261:13581–13587
- Kunkel TA (1990) Misalignment-mediated DNA synthesis errors. *Biochemistry* 29:8003–8011
- La Roche J, Snyder M, Cook DI, Fuller K, Zouros E (1990) Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. *Mol Biol Evol* 7:45–64
- Lai Y, Sun F (2003) The relationship between microsatellite slippage mutation rate and the number of repeat units. *Mol Biol Evol* 20:2123–2131
- Lavrov DV (2007) Key transitions in animal evolution: a mitochondrial DNA perspective. *Integr Comp Biol* 47:734–743
- Lessinger AC, Junqueira AC, Conte FF, Azeredo-Espin AM (2004) Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339:1–6
- Mabuchi K, Miya M, Satoh TP, Westneat MW, Nishida M (2004) Gene rearrangements and evolution of tRNA pseudogenes in the mitochondrial genome of the parrotfish (Teleostei: Perciformes: Scaridae). *J. Mol. Evol.* 59:287–297
- Madani M, Subbotin SA, Ward LJ, Li X, DeBoer SH (2010) Molecular characterization of Canadian populations of potato cyst nematodes, *Globodera rostochiensis* and *G. pallida* using ribosomal nuclear RNA and cytochrome b genes. *Can J Plant Pathol* 32:142–153
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9:387–402
- Milbury CA, Gaffney PM (2005) Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. *Mar Biotechnol (NY)* 7:697–712
- Mindell DP, Sorenson MD, Dimcheff DE (1998) An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol Biol Evol* 15:1568–1571
- Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* 18:269–292
- Nour SM, Lawrence JR, Zhu H, Swerhone GD, Welsh M, Welacky TW, Topp E (2003) Bacteria associated with cysts of the soybean cyst nematode (*Heterodera glycines*). *Appl Environ Microbiol* 69:607–615
- Rand DM, Harrison RG (1989) Molecular population genetics of mtDNA size variation in crickets. *Genetics* 121:551–569
- Riepsamen A, Blok V, Phillips M, Gibson T, Dowton M (2008) Poly(T) variation within mitochondrial protein-coding genes in *Globodera* (Nematoda: Heteroderidae). *J Mol Evol* 66:197–209
- Rosengarten RD, Sperling EA, Moreno MA, Leys SP, Dellaporta SL (2008) The mitochondrial genome of the hexactinellid sponge *Aphrocallistes vastus*: evidence for programmed translational frameshifting. *BMC Genomics* 9:33

- Russell RD, Beckenbach AT (2008) Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. *J Mol Evol* 67:682–695
- San Mauro D, Gower DJ, Zardoya R, Wilkinson M (2006) A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. *Mol Biol Evol* 23:227–234
- Shinde D, Lai Y, Sun F, Arnheim N (2003) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)<sub>n</sub> and (A/T)<sub>n</sub> microsatellites. *Nucl Acids Res* 31:974–980
- Simison WB, Lindberg DR, Boore JL (2006) Rolling circle amplification of metazoan mitochondrial genomes. *Mol Phylogenet Evol* 39:562–567
- Subbotin SA, Vierstraete A, De Ley P, Rowe J, Waeyenberge L, Moens M, Vanfleteren JR (2001) Phylogenetic relationships within the cyst-forming nematodes (Nematoda, Heteroderidae) based on analysis of sequences from the ITS regions of ribosomal DNA. *Mol Phylogenet Evol* 21:1–16
- Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Biol Evol* 13:510–524
- Tamas I, Wernegreen JJ, Nystedt B, Kauppinen SN, Darby AC, Gomez-Valero L, Lundin D, Poole AM, Andersson SG (2008) Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc Natl Acad Sci USA* 105:14934–14939
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl Acids Res* 22:4673–4680
- Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* 27:6008–6013
- Tsang WY, Lemire BD (2002) Mitochondrial genome content is regulated during nematode development. *Biochem Biophys Res Comm* 291:8–16
- Viguera E, Canceill D, Ehrlich SD (2001) In vitro replication slippage by DNA polymerases from thermophilic organisms. *J Mol Biol* 312:323–333
- Wagner LA, Weiss RB, Driscoll R, Dunn DS, Gesteland RF (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucl Acids Res* 18:3529–3535
- Zevelev CE, Moritz C, Heideman A, Strum RA (1991) Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*; Gekkonidae). *J Mol Evol* 33:431–441