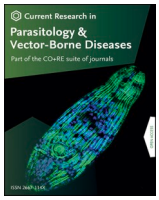
Current Research in Parasitology & Vector-Borne Diseases 9 (2026) 100346

Contents lists available at ScienceDirect

Current Research in Parasitology & Vector-Borne Diseases

journal homepage: www.sciencedirect.com/journal/current-research-in-parasitology-and-vector-borne-diseases



Long-read sequencing reveals that the mitochondrial genome of an individual *Ixodiphagus hookeri* is a mixture of structural variants with an invariable core region and heterogeneous repeat regions

Ron P. Dirks [a,*], Hans J. Jansen [a], Wouter J. Veneman [a], Jane Segobola [a], Valérie O. Baede [b], Veronika Blažeková [c], Michal Stanko [c,d], Bronislava Víchová [c], Hein Sprong [b]

[a] *Future Genomics Technologies B.V., Leiden, the Netherlands*
[b] *Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands*
[c] *Institute of Parasitology, Slovak Academy of Sciences (SAS), Košice, Slovakia*
[d] *Institute of Zoology, Slovak Academy of Sciences (SAS), Bratislava, Slovakia*

ARTICLE INFO

Keywords:
Nanopore sequencing
Long-read
Repeat region
Wasp
Tick

ABSTRACT

The encyrtid wasp *Ixodiphagus hookeri* (Hymenoptera: Encyrtidae) predates on a broad range of tick species, including the important disease vectors of the *Ixodes ricinus* complex. This wasp is a promising candidate for the biological control of tick populations. To support future studies on the biological interaction between ticks and tick wasps, the complete genome sequence of *I. hookeri* was determined using nanopore long-read technology. This work presents the complete mitochondrial genome of a singular *I. hookeri*. In-depth annotation of multiple nanopore long reads revealed that the mtDNA genome is a mixture of structural variants. It consists of an invariable ~15.2-kb core flanked by variable regions. Together, this variable region contains tens of copies of a 92-bp repeat unit, two copies of a ~340-bp repeat unit, and up to more than 1500 copies of a TA-dinucleotide. Parts of the mtDNA core sequence and the repeat region have integrated a few hundred times into the nuclear genome of *I. hookeri*. Our study shows that long-read DNA sequencing is essential for reliable *de novo* assembly of insect mtDNA genomes.

1. Introduction

Ixodiphagus is a genus of encyrtid wasps that predate on ticks. Since more than a century, people have been studying their potential use as biological agents to control ixodid tick populations and subsequently reduce the burden of tick-borne diseases (reviewed in Samish et al., 2004; Ramos et al., 2023). At least 15 species of the genus *Ixodiphagus* have been described (Ramos et al., 2023; Giannotta et al., 2024). Scientific interest is mostly focused on *Ixodiphagus hookeri* (Howard, 1908), the "tick wasp", which is found in Asia, Africa, North America and Europe (Samish et al., 2004; Krawczyk et al., 2020). Despite this strong interest, the complete mitochondrial genome sequence of *I. hookeri* has not yet been described. In June 2025, NCBI's nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore) contained about 150 entries of *Ixodiphagus*, including complete circular mitochondrial genomes of Australasian species *I. brunneus* ($n = 4$; 15,514–15,516 bp), *I. taiaroaensis* ($n = 1$; 15,470 bp) and some unknown species ($n = 3$; 14,950–15,056 bp) (Giannotta et al., 2024). The remaining entries consisted of partial sequences of the mitochondrial cytochrome *c* oxidase subunit 1 (*cox*1) gene (also from *I. hookeri*) and the 28S (large subunit) ribosomal RNA gene. To support future ecological, gene-expression, and evolutionary studies on the interaction between ticks and tick wasps, we sequenced the complete genome of *I. hookeri* using nanopore long-read technology. We discovered that its mitochondrial genome contains several features that were not found in the complete mitogenomes of the closely related Australasian species (Giannotta et al., 2024). In addition, we found a large number of nuclear mitochondrial DNA segments (NUMTs). NUMTs are pseudogenes derived from the translocation of mitochondrial sequences into the nuclear chromosomes (Lopez et al., 1994; recently reviewed by Xue et al., 2023) and are also found in insect genomes (reviewed by Hebert et al., 2023).

In this study, we present the in-depth analysis of the mitogenome of

* Corresponding author.
E-mail address: dirks@futuregenomics.tech (R.P. Dirks).

https://doi.org/10.1016/j.crpvbd.2026.100346
Received 21 August 2025; Received in revised form 19 December 2025; Accepted 1 January 2026
Available online 2 January 2026

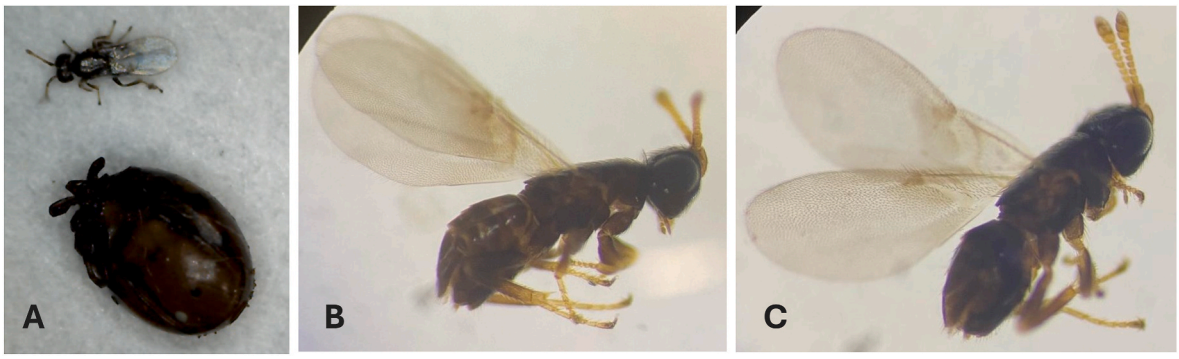


Fig. 1. Photographs of *Ixodiphagus hookeri* (all photos were taken by the authors). **A** Macroscopic photograph of the *I. hookeri* (top) next to the engorged *Ixodes ricinus* nymph (bottom) from which it emerged. **B**, **C** Microscopic photographs of *I. hookeri*.

I. hookeri and a genome-wide overview of NUMTs in the chromosomes of this wasp. Our research shows that long-read sequencing has become an essential tool for the reliable assembly of complete mitochondrial genomes.

2. Materials and methods

2.1. Source and morphological identification of specimens of Ixodiphagus hookeri

The *Ixodiphagus hookeri* wasp used in this study originated from an *Ixodes ricinus* nymph collected using the standard flagging method, employing a 1 × 1 m flag, in the Slovak Karst National Park (212 m above sea level; 48°34′N, 20°46′E) in late March 2020. In the laboratory, the nymph was placed on a mouse to blood-feed until fully engorged. It was then transferred to a climabox (Memmert SF 75, Germany), with temperature and relative humidity set to 22.5 °C and 90%, respectively. After 28 days, a wasp emerged from the tick (Fig. 1). The species and sex of the endoparasitoid wasp were identified based on morphological characteristics, as described by Gahan (1934). The specimen was stored at −80 °C and transported on dry ice for molecular analysis (in 2024).

2.2. DNA extraction and nanopore sequencing

DNA was extracted from an individual wasp specimen using DNeasy Blood & Tissue Kit (Qiagen Benelux BV, Venlo, The Netherlands) according to the manufacturer's instructions. The quality of the DNA was checked using Genomic DNA ScreenTape on an Agilent 4200 TapeStation System (Agilent Technologies Netherlands BV, Amstelveen, The Netherlands). Quantification of the DNA sample was performed using a Qubit 3.0 Fluorometer (Life Technologies Europe BV, Bleiswijk, The Netherlands) and revealed a total DNA yield of 72 ng. A nanopore sequencing library was prepared using the Ligation Sequencing Kit V14 (SQK-LSK114) according to the manufacturer's instructions (Oxford Nanopore Technologies (ONT), Oxford, UK). The library was run on an R10.4.1 PromethION flowcell (FLO-PRO114M; ONT, Oxford, UK) for ~45 h. Super-accuracy basecalling of raw POD5 data was done using Dorado v0.9.0 and resulted in 43,449,713 reads (62.9 Gb) with a read

Table 1
Characteristic structural features of high-quality nanopore reads derived from individual *Ixodiphagus hookeri* mtDNA molecules. Ten structural features could be discriminated (in clockwise direction): a ~15.2-kb core sequence, a TA-repeat, multiple 92-bp repeat units, a partial version (position 1–56) of the 92-bp repeat unit, Box_Q, another TA-repeat, Box_Q, a partial version (position 1–56) of the 92-bp repeat unit, multiple 92-bp repeat units, and a partial version (position 61–92) of the 92-bp repeat unit.

Read category	Unique identifier	Size (nt)	Box_Q reverse (bp)	rep_1–56-reverse (# copies)	rep_1–92-reverse (# copies)	rep_61–92-reverse (# copies)	mtDNA core (bp)	[TA] (# copies)	rep_1–92-forward (# copies)	rep_1–56-forward (# copies)	Box_Q forward (bp)	[TA] (# copies)
Full-length core region	03fa4618	16,464	n.a.	n.a.	1p	1	15,163	15	12 + 1p	n.a.	n.a.	n.a.
	23425c59	19,782	n.a.	n.a.	4 + 1p	1	15,155	18	44 + 1p	n.a.	n.a.	n.a.
	b547a18f	19,020	n.a.	n.a.	24 + 1p	1	15,157	19	16 + 1p	n.a.	n.a.	n.a.
	29ac7f52	19,783	n.a.	n.a.	18 + 1p	1	15,173	17	30 + 1p	n.a.	n.a.	n.a.
	dc777027	20,412	n.a.	n.a.	22 + 1p	1	15,159	17	33 + 1p	n.a.	n.a.	n.a.
	08c36f79	17,043	n.a.	n.a.	16 + 1p	1	15,150	17	3 + 1p	n.a.	n.a.	n.a.
	f9d08e3f	16,843	n.a.	n.a.	1p	1	15,125	16	17 + 1p	n.a.	n.a.	n.a.
	880bc874	20,451	n.a.	n.a.	7 + 1p	1	15,040	17	48	1	167 (p)	n.a.
Full-length repeat region	821384 cb	23,348	344	1	25	1	7621 + 7550	21	37	1	337	780
	b66eb5c6	19,062	344	1	28	1	6819 + 4262	17	48	1	336	43
5′ core and repeat region	4042742c	8532	344	1	28	1	4986	n.a.	n.a.	n.a.	n.a.	241
	575e7b8d	10,124	344	1	28	1	5932	n.a.	n.a.	n.a.	n.a.	575
	e3710536	10,415	344	1	34	1	4214	n.a.	n.a.	n.a.	n.a.	1306
	65af9bd9	14,483	342	1	28	1	11,020	n.a.	n.a.	n.a.	15 (p)	199
	7b7386ad	9044	341	1	28	1	5695	n.a.	n.a.	n.a.	n.a.	159
3′ core and repeat region	ae5f0363	9368	344	1	5 + 1p	n.a.	1124	17	38	1	336	1689
	6a372dec	7839	n.a.	n.a.	n.a.	n.a.	1836	17	48	1	336	565
	7e6dd5b5	10,420	336	1	45 + 1p	n.a.	963	17	48	1	336	28
	7291e499	17,359	333	1	30 + 1p	n.a.	10,116	17	37	1	335	75
	908b4e69	7067	335	1	1p	n.a.	1495	15	48	1	338	114
	0e298127	12,876	n.a.	n.a.	n.a.	n.a.	7701	18	48	1	336	154
	63508429	8123	n.a.	n.a.	n.a.	n.a.	3242	18	48	1	338	n.a.

Abbreviations: bp, number of basepairs; n.a., not applicable; p, partial.

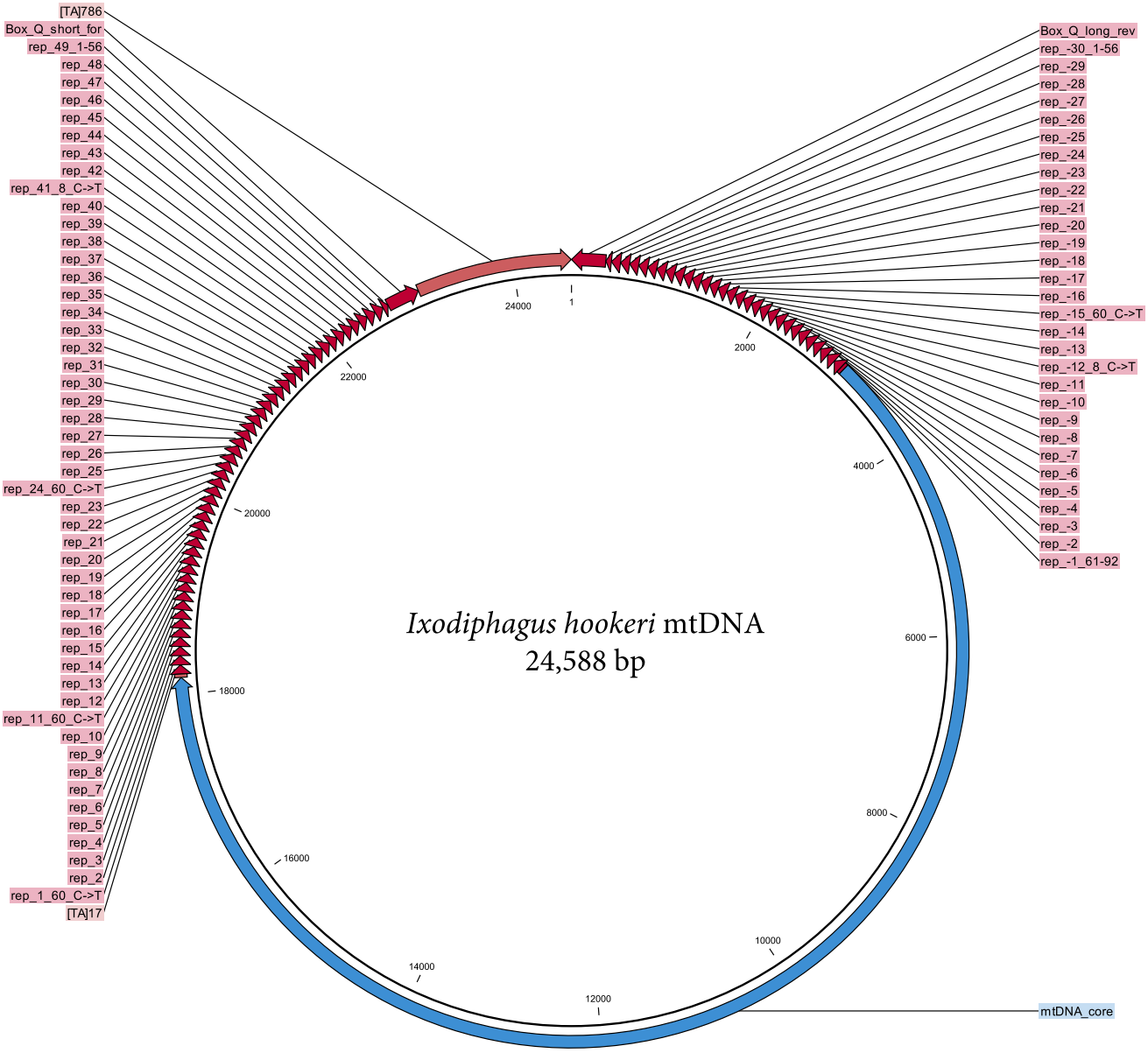


Fig. 2. Schematic representation of the consensus features of the complete mitochondrial genome of *Ixodiphagus hookeri*, consisting of a 15,166-bp core sequence (*blue*), and a 9422-bp repeat sequence (*red/pink*). In clockwise direction: (i) 15,166-bp core sequence (*blue*); (ii) 17 copies of a TA-dinucleotide (*pink*); (iii) 48 forward copies of the 92-bp repeat unit (*red*), with a C→T transition at position 60 in repeat numbers +1, +11, +24, and with a C→T transition at position 8 in repeat number +41; (iv) one partial forward copy (bp 1–56) of the 92-bp repeat unit (*red*); (v) one forward copy of the long version of Box_Q (*red*); (vi) 786 copies of a TA-dinucleotide (*pink*); (vii) one reverse copy of the long version of Box_Q (*red*); (viii) one partial reverse copy (bp 1–56) of the 92-bp repeat unit (*red*); (ix) 28 reverse copies of the 92-bp repeat unit (*red*), with a C→T transition at position 60 in repeat number −15 and with a C→T transition at position 8 in repeat number −12; and (x) one partial reverse copy (bp 61–92) of the 92-bp repeat unit (*red*).

N50 value of 4393 nt.

2.3. Identification and annotation of the mitochondrial genome sequence

The complete genome sequence of *I. hookeri* was *de novo* assembled using Hifiasm v.0.23.0-r691 (Cheng et al., 2021) and will be described in more detail elsewhere. The total assembly size was ~367 Mb, divided over 538 contigs with an N50 contig length of ~25 Mb. The assembled genome sequence had 99.2% complete BUSCOs (Arthropoda, $n = 1013$). A BLASTn search (CLC Main Workbench v25.0) with mitochondrial genome sequences of Australian *Ixodiphagus* species (Giannotta et al., 2024; GenBank: PP134637-PP134646) against the *I. hookeri* genome assembly resulted in multiple hits: circular contig_000186 (44,455 bp) and multiple linear contigs with a minimum size of 107 kb. Alignment of nanopore reads against mtDNA reference sequences was done using Minimap2 v2.27 and filtered reads were obtained using Samtools v1.18. Reads were manually annotated using CLC Main Workbench v25.0.1. The core sequence of the mitochondrial genome of *I. hookeri* was

			8C→T	60C→T		8C→T		60C→T							8C→T			
Read category	Unique identifier	Size (nt)	rep -33	rep -15	rep -13	rep -12	rep -10	rep +1	rep +9	rep +11	rep +14	rep +16	rep +22	rep +24	rep +30	rep +31	rep +35	rep +41
Full-length core region	03fa4618	16,464						T	C	T								
	23425c59	19,782						T	C	T	C	C	C	T	C	C	C	T
	b547a18f	19,020		T	C	T	C	T	C	T	C	C						
	29ac7f52	19,783		T	C	T	C	T	T	C	C	C	T	C	C			
	dc777027	20,412		T	C	T	C	T	C	T	C	C	C	T	C	C		
	08c36f79	17,043		T	C	T	C	T										
	f9d08e3f	16,843						T	C	T	C	T						
	880bc874	20,451						T	C	T	C	C	C	T	C	C	C	T
Full-length repeat region	821384cb	23,348		C	T	C	T	T	C	C	C	C	C	C	C	C	T	
	b66eb5c6	19,062		T	C	T	C	T	C	T	C	C	C	T	C	C	C	T
5' core and repeat region	4042742c	8,532		T	C	T	C											
	575e7b8d	10,124		T	C	T	C											
	e3710536	10,415	T	C	T	C	T											
	65af9bd9	14,483		T	C	T	C											
	7b7386ad	9,044		T	C	T	C											
3' core and repeat region	ae5f0363	9,368						T	C	C	T	C	C	C	C	T	C	
	6a372dec	7,839						T	C	T	C	C	C	T	C	C	C	T
	7e6dd5b5	10,420						T	C	T	C	C	C	T	C	Δ2nt	C	T
	7291e499	17,359						T	C	T	C	C	C	C	T	C	C	
	908b4e69	7,067						T	C	T	C	C	C	T	C	C	C	Δ2nt
	0e298127	12,876						T	C	T	C	C	C	T	C	C	C	T
	63508429	8,123						T	C	T	C	C	C	T	C	C	C	T

Fig. 3. Sequence variation at positions 8 and 60 of the 92-bp repeat unit. C nucleotides are shown in blue, T nucleotides are shown in green, grey boxes indicate that the repeat unit is absent, and orange boxes indicate a 2-nucleotide deletion at position 7–8.

annotated using MITOS2 (Donath et al., 2019) *via* the Galaxy webserver (https://usegalaxy.org) and according to the invertebrate mitochondrial code (translation table 5). Five 1-nt indels in homopolymer stretches of more than ten nucleotides were manually corrected based on open reading frames in the complete mtDNA sequences of related species of the family Encyrtidae. The MITOS2-based annotation of both ribosomal RNA genes and of the *nad*6 gene was incomplete and therefore corrected based on the annotation of mtDNAs of closely related species.

2.4. Phylogenetic analysis

To examine the phylogeny of *I. hookeri* in the family Encyrtidae, complete nucleotide sequences of all 13 protein-coding genes were extracted from the annotated mitochondrial genomes of 9 representative encyrtid wasps (Giannotta et al., 2024) and translated according to the invertebrate mitochondrial code (translation table 5). Nucleotide and amino-acid sequences were concatenated and then used in multiple sequence alignments. The aligned sequences were subsequently used according to the Maximum Likelihood method (construction method: Neighbor Joining; substitution model: Jukes Cantor (CDSs); bootstrapping: 1000 replicates). All tools were part of CLC Main Workbench v25.0.1.

2.5. Mapping of nuclear mitochondrial DNA (NUMT) segments

Sequences of all 13 protein-coding genes (PCGs) and both rRNA genes were extracted from the annotated *I. hookeri* mtDNA sequence and used in separate BLASTn searches against the assembled *I. hookeri* genome (E-value = 1E-10). All corresponding sequences in chromosomal contigs were manually annotated (CLC Main Workbench v25.0.1).

3. Results

3.1. Structure of the I. hookeri mtDNA molecule

The *de novo* assembled Hifiasm contigs ptg_000186 (44,455 bp) and ptg_000177 (107,000 bp) contained multiple full-length and partial copies of a 15.2-kb mtDNA core sequence, tens of copies of a 92-bp repeat unit, several copies of a 0.34-kb repeat unit (which we named Box_Q) and hundreds of copies of a TA-dinucleotide repeat (data not shown). Since both contigs looked like assembly artefacts, we decided to determine the exact structure of the mitochondrial DNA *via* in-depth analysis of individual nanopore reads. The overall structure of the *I. hookeri* mtDNA molecule could be deduced *via* manual annotation of 22 overlapping reads (Table 1, Fig. 2, and Supplementary Fig. S1). Eight reads had a complete and uninterrupted copy of the 15.2-kb mtDNA core, flanked by partial repeat regions. Five reads contained the 5′ part of the mtDNA core plus a large part of the 5′ flanking repeat region. Seven reads contained the 3′ part of the mtDNA core plus a large part of the 3′ flanking repeat region. Two reads were particularly informative because they bridged the complete repeat region: read_821384 cb (23,348 bp) was derived from a full-length circular mtDNA molecule that was linearized in the middle of the core sequence, and read_b66eb5c6 was nearly complete and only lacked ~4.1 kb of the central mtDNA core sequence. Ten structural features of the *I. hookeri* mtDNA genome could be discriminated (in clockwise direction): a 15.2-kb core sequence, 15–21 copies of a TA-dinucleotide, 37 or 48 forward copies of the 92-bp repeat unit, one partial forward copy (position 1–56) of the 92-bp repeat unit, one forward copy of Box_Q, between 28 and 1689 copies of a TA-dinucleotide, one reverse copy of Box_Q, one partial reverse copy (position 1–56) of the 92-bp repeat unit, up to 46 reverse copies (mostly 28 copies) of the complete 92-bp repeat unit, and one partial reverse copy (position 61–92) of the 92-bp repeat unit. Numbering of the forward

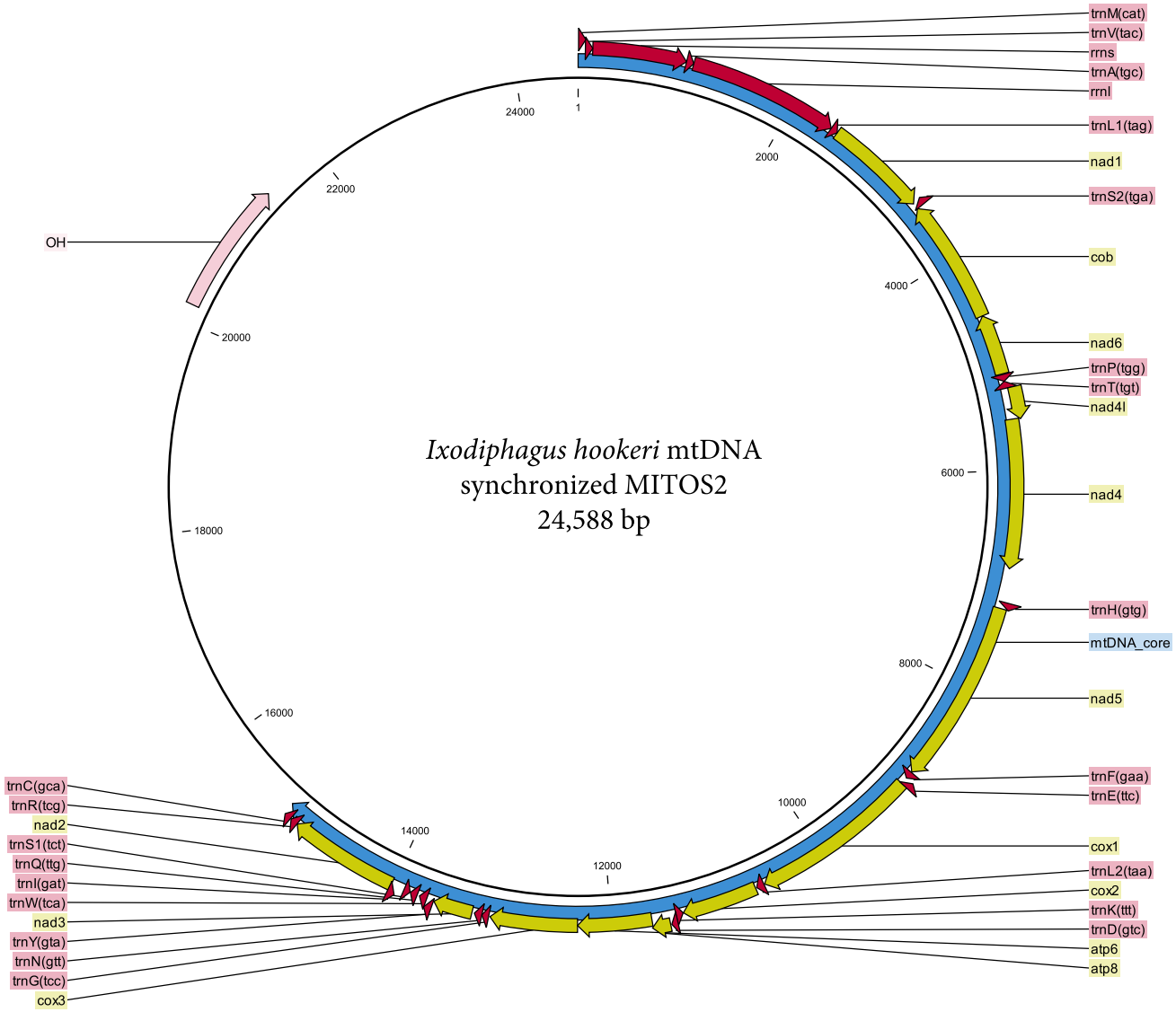


Fig. 4. MITOS2-based annotated features in the *Ixodiphagus hookeri* mtDNA. The mtDNA core region is shown in *blue*. The protein-coding genes ($n = 13$) are shown in *yellow*. The tRNA genes ($n = 22$) are shown in *red* and named after their corresponding 1-letter amino-acid code. The small (rrns) and large (rrnl) ribosomal RNA genes are also shown in *red*. OH: the predicted origin of heavy strand replication, which coincides with the [TA]786 repeat in the repeat region.

copies of the 92-bp repeat unit started at the first repeat unit (rep_1) located immediately downstream of the mtDNA core sequence. Numbering of the reverse copies of the 92-bp repeat unit started at the partial first repeat unit (rep_-1_61–92) located immediately upstream of the mtDNA core sequence, and those numbers were preceded by a minus sign.

Multiple sequence alignment of 895 copies of the 92-bp repeat units with numbered locations, derived from the 22 annotated nanopore reads, revealed that 86% of the repeat units was 100% identical to the consensus sequence 5′-TAATCCC**C**TC-AATATTACAT-ATAATTAAAT-TTTTAAAAGT-AAAATTTTTA-ATTTTACCA**C**-TTATTTTTAG-TTATTTT-GAC-AAAATTATAT-CG-3′ (Supplementary Fig. S2). Sequence variation was mainly detected at two positions of the repeat sequence and at very specific repeat locations (Fig. 3 and Supplementary Fig. S3): at nucleotide position 8, a T instead of a C was found in the majority of repeat numbers −12 (82%) and +41 (88%), whereas at nucleotide position 60, a T instead of a C was found in the majority of repeat numbers −15 (82%), +1 (100%), +11 (81%) and +24 (69%). Read_821384 cb and read_e3710536 showed a T at position 8 in repeat number −10 instead of number −12 and a T at position 60 in repeat number −13 instead of number −15 (Fig. 3), which suggests deletion of two repeat units in those two reads compared to the other reads. Similarly, read_29ac7f52 showed a T at position 60 in repeat numbers +9 and +22 instead of, respectively, numbers +11 and +24, which also suggests deletion of two repeat units in this read compared to the other reads. Multiple sequence alignment of 20 full-length copies of Box_Q (Supplementary Fig. S4) showed that it has a 344-bp variant (Box_Q_long) and a 336-bp variant (Box_Q_short). In the 22 manually annotated nanopore reads, 100% of the forward copies of Box_Q was the short variant (9 out of 9), whereas 73% of the reverse copies was the long variant (8 out of 11), and 27% was the short variant (3 out of 11). In summary, the mitochondrial genome of a singular *I. hookeri* wasp is a mixture of multiple structural variants with the following ten consecutive consensus features (Fig. 2): (i) a 15,166-bp core sequence, and a 9422-bp repeat sequence, which

Table 2
Characteristics of the mitochondrial genome of *I. hookeri*.

Feature	Feature type	Protein name	Strand	Start sites	Stop sites	Size (bp)	Anticodon	Start codon	Stop codon
trnM	tRNA		plus	5	68	64	cat		
trnV	tRNA		plus	66	131	66	tac		
rrns	rRNA		plus	132	977	846			
trnA	tRNA		plus	978	1044	67	tgc		
rrnl	rRNA		plus	1045	2405	1361			
trnL1	tRNA		plus	2406	2471	66	tag		
nad1	PCG	ND1	plus	2472	3404	933		ATA	TAG
trnS2	tRNA		minus	3395	3460	66	tga		
cob	PCG	CYTB	minus	3460	4581	1122		ATG	TAA
nad6	PCG	ND6	minus	4583	5128	546		ATG	TAA
trnP	tRNA		plus	5135	5196	62	tgg		
trnT	tRNA		minus	5197	5260	64	tgt		
nad4l	PCG	ND4L	plus	5261	5548	288		ATT	TAA
nad4	PCG	ND4	plus	5542	6882	1341		ATG	TAA
trnH	tRNA		plus	7184	7248	65	gtg		
nad5	PCG	ND5	plus	7246	8916	1671		ATT	TAA
trnF	tRNA		plus	8922	8988	67	gaa		
trnE	tRNA		minus	8989	9053	65	ttc		
cox1	PCG	COI	plus	9052	10,578	1527		ATG	TAA
trnL2	tRNA		plus	10,585	10,655	71	taa		
cox2	PCG	COII	plus	10,667	11,347	681		ATT	TAA
trnK	tRNA		minus	11,350	11,415	66	ttt		
trnD	tRNA		plus	11,414	11,469	56	gtc		
atp8	PCG	ATP8	plus	11,480	11,641	162		ATT	TAA
atp6	PCG	ATP6	plus	11,635	12,300	666		ATG	TAA
cox3	PCG	COIII	plus	12,301	13,086	786		ATG	TAA
trnG	tRNA		plus	13,099	13,160	62	tcc		
trnN	tRNA		plus	13,161	13,229	69	gtt		
nad3	PCG	ND3	plus	13,252	13,605	354		ATA	TAA
trnY	tRNA		plus	13,610	13,674	65	gta		
trnW	tRNA		plus	13,674	13,742	69	tca		
trnQ	tRNA		plus	13,765	13,829	65	ttg		
trnS1	tRNA		minus	13,833	13,891	59	tct		
trnI	tRNA		minus	13,954	14,021	68	gat		
nad2	PCG	ND2	plus	14,014	15,012	999		ATT	TAA
trnR	tRNA		plus	15,015	15,083	69	tcg		
trnC	tRNA		plus	15,090	15,154	65	gca		
Repeat region	Repeat region		plus	15,167	24,588	9422			
OH	Origin of replication		plus	20,173	21,411	1239			

Abbreviations: OH, origin of heavy strand replication; PCG, protein-coding gene; rrnl, large 16S ribosomal RNA; rrns, small 12S ribosomal RNA; trn*, transfer RNA plus 1-letter amino-acid code.

consists of (ii) 17 copies of a TA-dinucleotide, (iii) 48 forward copies of the 92-bp repeat unit, with a C→T transition at position 60 in repeat numbers +1, +11, +24, and with a C→T transition at position 8 in repeat number +41, (iv) one partial forward copy (bp 1–56) of the 92-bp repeat unit, (v) one forward copy of the long version of Box_Q, (vi) 768 copies of a TA-dinucleotide (calculated by Hifiasm), (vii) one reverse copy of the long version of Box_Q, (viii) one partial reverse copy (bp 1–56) of the 92-bp repeat unit, (ix) 28 reverse copies of the 92-bp repeat unit, with a C→T transition at position 60 in repeat number −15 and with a C→T transition at position 8 in repeat number −12, (x) one partial reverse copy (bp 61–92) of the 92-bp repeat unit. The A + T content of the complete ~24.6-kb *I. hookeri* mtDNA is 87.0%, divided over 86.9% in the ~15.2-kb core and 87.3% in the ~9.4-kb repeat region.

3.2. Functional annotation and phylogenetic analysis of the I. hookeri mtDNA molecule

The starting point of the mtDNA genome sequence was moved to the first nucleotide of the core sequence, which is immediately downstream of the partial reverse copy of the 92-bp repeat unit (structural feature 10 in Fig. 2) and then the mtDNA genome was annotated using MITOS2. The mtDNA core of *I. hookeri* contains 13 PCGs, 2 ribosomal RNA genes (rRNAs), and 22 tRNA genes (tRNAs) (Fig. 4, Table 2). All PCGs follow the invertebrate mitochondrial code (translation table 5). Genes coding for ND1 (*nad*1) and ND3 (*nad*3) start with an ATA codon, genes coding for ATP6 (*atp*6), CYTB (*cob*), COI (*cox*1), COIII (*cox*3), ND4 (*nad*4) and ND6 (*nad*6) start with an ATG codon, and genes coding for ATP8 (*atp*8), COII (*cox*2), ND2 (*nad*2), ND4L (*nad*4l), ND5 (*nad*5) and ND6 (*nad*6) start with an ATT codon. Twelve PCGs use TAA as a stop codon, and only the gene coding for ND1 (*nad*1) uses a TAG stop codon. The TGA codon codes for tryptophan instead of for a stop, ATA codes for methionine instead of for isoleucine, and AGA and AGG code for serine instead of for arginine (all according to the invertebrate mitochondrial translation table 5). The tRNA genes vary in size from 56 bp (*trnD*) to 71 bp (*trnL2*). All predicted tRNAs show the expected 2D-structure, except that trnD lacks the T-arm and trnS1 lacks the D-arm (Supplementary Fig. S5). MITOS2-based predictions of both rRNA genes were partial, and therefore the rRNA genes were manually deduced from the annotation of the mtDNA sequence of a closely related encyrtid wasp (*Ooencyrtus plautus*; GenBank: OP442361; Xing et al., 2022).

A detailed phylogenetic analysis of various Australasian members of the genus *Ixodiphagus* was recently published (Giannotta et al., 2024). We used a representative selection of nine encyrtid wasp mtDNA sequences from this study to generate maximum likelihood (ML) phylogenetic trees based on concatenated nucleotide and amino-acid sequences of the 13 PCGs (Supplementary Fig. S6 and S7; protein-based tree shown in Fig. 5). Both trees strongly support that *I. hookeri* forms a single clade with the Australasian members of the genus *Ixodiphagus* and is also closely related to *O. plautus*. This is further supported by analysis of the gene order in the mitogenomes of six representative species of encyrtid wasps (Fig. 6). The gene order is identical in all three members of the genus *Ixodiphagus*, whereas gene rearrangements are visible in the mitogenomes of the other encyrtid wasp species.

Table 3
Genome-wide overview of nuclear mitochondrial DNA segments (NUMTs) in *I. hookeri*. The actual sizes (in base pairs) of the original mitochondrial genes are shown in parentheses in the top row. The numbers in parentheses in the first column indicate the contig sizes (in base pairs). The numbers in each cell indicate how many times a gene was found in a contig and the numbers in parentheses indicate the sizes of the BLASTn hits (in base pairs).

Contig #	*rrn*S (864)	*rrn*L (1361)	*nad*1 (933)	*cob* (1122)	*nad*6 (510)	*nad*4L (288)	*nad*4 (1341)	*nad*5 (1671)	*cox*1 (1527)	*cox*2 (681)	*atp*8 (162)	*atp*6 (666)	*cox*3 (786)	*nad*3 (354)	*nad*2 (999)	Total	rep (92)
ptg000006l (45,474,516)	5 (129–194)	2 (481–962)	3 (249–741)	3 (231–1025)	1 (79)	1 (139)	7 (159–618)	6 (119–455)	2 (433–1253)	1 (206)	3 (79–107)	2 (64–92)	1 (245)	2 (74–106)	6 (120–620)	45	350
ptg000029l (36,154,831)	6 (97–615)	7 (122–800)	8 (81–873)	4 (118–298)	6 (41–535)	2 (131–158)	8 (144–518)	7 (79–988)	7 (52–552)	4 (72–537)	3 (89–133)	4 (130–385)	4 (61–583)	2 (260–342)	7 (147–1004)	79	4
ptg000014l (28,211,289)	3 (145–244)	12 (76–552)	7 (110–900)	6 (81–839)	5 (104–544)	1 (298)	5 (76–464)	9 (84–368)	9 (97–902)	2 (166–681)	3 (78–158)	2 (356–452)	4 (144–687)	2 (74–86)	5 (950–585)	75	0
ptg000027l (27,738,704)	4 (145–347)	4 (248–1148)	5 (145–794)	5 (47–722)	3 (397–583)	6 (77–263)	6 (139–707)	4 (227–1106)	6 (35–1537)	2 (680–684)	5 (95–180)	7 (72–666)	6 (90–784)	4 (73–354)	5 (73–776)	72	6
ptg000035l (26,509,353)	4 (127–722)	2 (313–641)	1 (411)	4 (82–446)	1 (51)	3 (80–243)	4 (143–374)	6 (84–548)	3 (242–592)	4 (102–321)	0	2 (168–368)	2 (642–673)	2 (68–303)	3 (224–543)	41	54
ptg000009l (24,815,080)	1 (86)	1 (776)	3 (55–159)	4 (75–863)	2 (353–441)	4 (67–191)	10 (86–491)	5 (79–925)	7 (46–682)	2 (110–197)	1 (111)	3 (224–282)	1 (433)	1 (105)	1 (129)	46	3
ptg000018l (22,701,649)	0	0	0	3 (165–203)	2 (74–467)	2 (51–82)	1 (169)	0	4 (318–995)	1 (611)	4 (77–141)	4 (123–292)	2 (106–135)	2 (282–325)	0	25	5
ptg000016l (21,760,137)	4 (97–339)	2 (214–506)	2 (108–145)	0	2 (143–411)	1 (288)	2 (473–524)	3 (384–687)	1 (698)	2 (501–599)	1 (155)	1 (73)	0	0	3 (170–522)	24	161
ptg000023l (18,260,898)	3 (122–276)	0	0	3 (127–176)	4 (74–421)	3 (173–263)	3 (194–1174)	2 (111–1635)	4 (88–261)	2 (105–670)	2 (100–160)	2 (289–293)	1 (379)	2 (282–325)	1 (471)	32	9
ptg000028l (18,250,185)	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (79)	0	1	
ptg000024l (17,383,890)	1 (143)	0	0	0	0	0	0	0	0	0	0	0	0	1 (118)	0	2	
ptg000010l (11,265,424)	0	0	3 (154–239)	0	0	0	0	0	0	0	1 (153)	1 (666)	0	0	2 (130–258)	7	0
ptg000015l (8,774,455)	1 (154)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
ptg000025l (7,867,900)	1 (743)	1 (74)	1 (292)	0	1 (189)	2 (207–320)	1 (822)	2 (160–193)	1 (259)	0	0	2 (49–273	1 (393)	0	3 (206–903)	16	
ptg000047l (696,712)	0	0	0	0	1 (191)	0	0	0	0	0	0	0	0	0	0	1	
Total	33	31	33	32	28	25	47	44	44	20	23	30	22	19	36	469	592

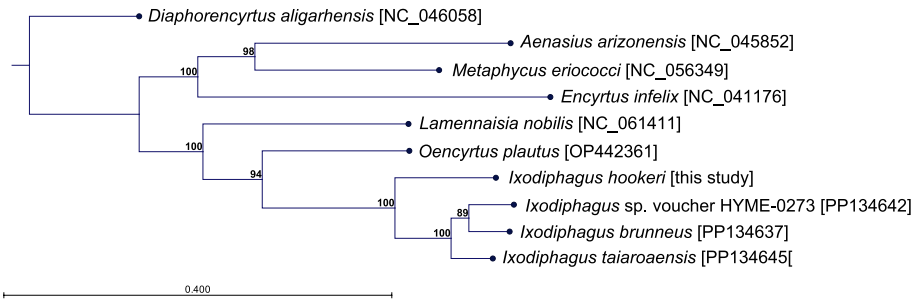


Fig. 5. Maximum likelihood phylogenetic tree of concatenated AAs extracted from the annotated mtDNA genomes of representative encyrtid wasp species. Bootstrap values (%) are based on 1000 replicates and shown above each root.

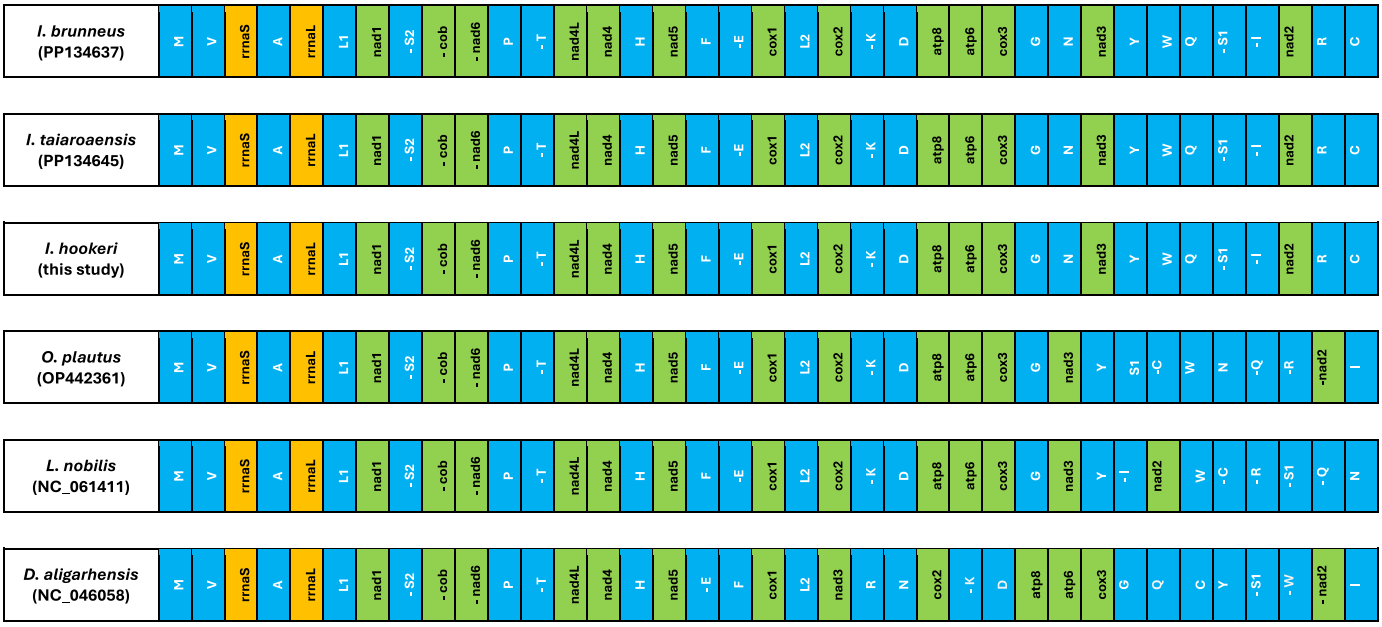


Fig. 6. Gene order in mitochondrial genomes of members of the family Encyrtidae. tRNA genes are shown in *blue*, ribosomal RNA genes are shown in *yellow*, and protein-coding genes are shown in *green*. tRNA genes are named after their corresponding 1-letter amino-acid code. Reverse gene orientation is indicated by a minus sign.

3.3. Genome-wide analysis of nuclear mitochondrial DNA sequences (NUMTs) in I. hookeri

When the consensus full-length ~24.6-kb *I. hookeri* mtDNA sequence was used in a BLASTn search against the *de novo* assembled *I. hookeri* genome sequence, we did not only find 100% identical hits in the two incorrectly assembled mitochondrial contigs (ptg_000186 and ptg_000177), but also multiple partial hits with 70–95% sequence identity in the 14 longest chromosomal contigs. To map these so-called nuclear mitochondrial DNA segments (NUMTs) in more detail, both rRNA genes, the 13 PCGs, and the repeat region were used in individual BLASTn searches and the corresponding hits were then used to create a genome-wide map of NUMTs in *I. hookeri* (Table 3 shows a complete summary of all the NUMTs; the most pronounced NUMTs are shown in Supplementary Fig. S8). A total of 469 genes divided over 198 independent NUMTs could be discriminated. More than 90% of the insertions were small and contained only one ($n = 120$), two ($n = 32$), three ($n = 15$), four ($n = 11$) or five ($n = 5$) PCGs and/or rRNA genes. Only eight insertions contained between 6 and 12 PCGs and/or rRNA genes and none of them contained the complete complement of all PCGs and rRNA genes. Seven insertions contained only mitochondrial repeat sequences. Less than 3% of the genes in the NUMTs were full length and the smallest recognizable parts of the genes were only about 40 nucleotides long (Table 3).

4. Discussion

As part of the *I. hookeri* genome project, we performed an in-depth analysis of its mitochondrial genome. Long-read genome assembly software tools (e.g. Canu, Flye, Hifiasm) have trouble with *de novo* assembly of small circular genomes, such as plasmids and mitochondrial genomes (Johnson et al., 2023). This may be caused by the higher copy number of these small circular genomes compared to that of the nuclear genome. Proper automated assembly of the mitochondrial genome sequence of *I. hookeri* was further complicated by a high number of (near) identical tandem repeats in the ~9.4-kb repeat region and by our observation that the mitochondrial genome of a singular *I. hookeri* specimen is not a homogeneous population of molecules, but rather a mixture of multiple structural variants. This was discovered *via* careful manual annotation of tens of individual long nanopore reads. Variation was found at multiple sites within the long repeat region, namely in the size of the TA-repeat region immediately downstream of the core, in the number of 92-bp repeat units, in the size of Box_Q (short or long) and especially in the size of the TA-repeat region between both copies of Box_Q. It is not known whether these mtDNA variants are already present in the wasp's egg or whether the mitochondrial genome starts out as a homogeneous population of molecules, which later undergoes intermolecular and/or intramolecular recombination, resulting in deletion, expansion, and/or inversion of repeat units. Since the size of the

individual 92-bp repeat units in each nanopore read is always constant, the variation in the number of repeat units cannot result from sequencing and/or base-calling errors. We also re-base-called the raw POD5 nanopore dataset using an alternative Dorado base-calling model that recognizes 4-methylcytosine, 5-methylcytosine, 5-hydroxymethylcytosine, and N6-methyladenosine and did not find significant evidence for any of these modified bases, which strongly suggests that the C→T transitions are actual mutations, rather than base modifications (data not shown). We can also rule out that some of the mtDNA variation corresponds to NUMTs, because all NUMT-derived nanopore reads show multiple mutations and indels and can be easily discriminated from the actual mtDNA-derived reads.

The mean read length of the nanopore reads was 1.5 kb, and the read N50 was 4.4 kb, which made it challenging to find full-length mtDNA reads. Indeed, only 2 out of a total of more than 44 million reads bridged the complete repeat region of ~9.4 kb, thereby connecting both ends of the mtDNA core and providing direct evidence for a circular mitochondrial genome. The nanopore sequencing library was made according to ONT's DNA ligation protocol. This means that sequencing adapters are ligated to linear DNA molecules and that circular mtDNA can only be sequenced after linearization due to physical shearing or DNase activity. We also applied shallow sequencing to DNA of another *I. hookeri* specimen that was processed using ONT's transposase-based rapid sequencing protocol, which should not discriminate between linear and circular DNA molecules; however, the percentage of mitochondrial reads in the transposase-based library was not higher than that in the ligase-based library (data not shown). This could mean that a considerable part of the intracellular mtDNA is linear. Alternatively, many mtDNA molecules may have been damaged during sample storage or DNA extraction, although this would contrast with the very high N50 contig length (~25 Mb) of the nuclear genome assembly. Therefore, it remains uncertain whether the mtDNA of *I. hookeri* is only circular or a mixture of circular and linear molecules.

An annotation of the mtDNA core shows that it contains the typical set of 13 PCGs, 22 tRNA and 2 rRNA genes. The translation of the PCGs adheres to the invertebrate codon table 5. Phylogenetic analysis confirms the close evolutionary link between *I. hookeri* and the Australasian *Ixodiphagus* family members; however, the published mtDNA genome sequences of the Australasian *Ixodiphagus* family members are much smaller than that of *I. hookeri* and lack the long repeat region (Giannotta et al., 2024). Since the Australasian *Ixodiphagus* mtDNA sequences were *de novo* assembled from Illumina short reads, the long repeat region may have collapsed into a much shorter version. The ~400-bp long presumed control region (CR) between the trnM and trnC genes in the mtDNA of *I. brunneus* (GenBank: PP134637, PP134638, PP134639, and PP134646) contains two perfect copies and one partial copy of a ~74-bp repeat unit. Distantly related copies of another ~74-bp repeat unit are present in the presumed 400-bp CR of the mtDNA of *I. taiaroaensis* (GenBank: PP134645). However, the sequences of these ~74-bp repeat units are not homologous with that of the 92-bp repeat in the mtDNA of *I. hookeri*. The presence of a long repeat region in the mtDNA of an insect is not a unique feature of *I. hookeri*. The mtDNA sequence of *Nasonia vitripennis* contains a 7.7-kb GC-rich repeat region with 29 copies of a 217-bp repeat unit (Lin et al., 2021). The importance of long-read sequencing technologies for proper assembly of mitochondrial genome sequences was further demonstrated by Kinkar et al. (2021), when they showed that the mtDNA genome of the carcinogenic parasitic worm *Schistosoma haematobium* has a long and variable repeat region. Even for vertebrate mitochondrial genomes, it is becoming increasingly clear that essential details are missed if only short-read sequencing methods are used (Formenti et al., 2021).

In our study, we identified about two hundred NUMTs. This is in agreement with a recent study on the frequency of NUMTs in insect genomes (Hebert et al., 2023). The abundant presence of mitochondrial sequences in the nuclear genome could further complicate the proper *de novo* assembly of the mtDNA genome. Two of the NUMT loci identified in our study contained more than a hundred copies of the 92-bp repeat unit, which would strongly interfere with an Illumina short-read-based mtDNA genome assembly. In conclusion, reliable *de novo* assembly of insect mtDNA genomes requires long-read DNA sequencing and may even require in-depth analysis of individual long reads.

5. Conclusions

We discovered that the mitochondrial genome of the "tick wasp" *I. hookeri* consists of an invariable ~15.2-kb core region, flanked by a complex repeat region. The repeat region is heterogeneous and contains a variable number of a 92-bp repeat unit, two copies of a ~340-bp repeat unit and hundreds of copies of a TA-dinucleotide repeat unit. In addition, the chromosomes of *I. hookeri* contain a few hundred NUMTs, with 70–95% sequence identity to the actual mitochondrial genome. The complex and heterogeneous nature of the wasp's mitochondrial genome, combined with the presence of highly homologous NUMTs causes challenges for *de novo* assembly of the complete mitogenome from Illumina short reads, and also for modern long-read assembly tools, such as Flye and Hifiasm. In this study, we show that manual in-depth annotation of multiple high-quality nanopore long reads is a good strategy for unravelling the true nature of the complex mitochondrial genome of *I. hookeri*.

Ethical approval

All animal experiments were carried out in accordance with the Animal Use Protocol approved by the ethical committee of the Biomedical Research Centre of the Slovak Academy of Sciences and the State Veterinary and Food Administration of the Slovak Republic (facility number SKP40011/SKU01016, permit number 2725/15–221/292/16-221k).

CRediT authorship contribution statement

Ron P. Dirks: Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Hans J. Jansen:** Data curation, Investigation, Writing – review & editing. **Wouter J. Veneman:** Investigation, Methodology, Writing – review & editing. **Jane Segobola:** Investigation, Methodology, Writing – review & editing. **Valérie O. Baede:** Investigation, Writing – review & editing. **Veronika Blažeková:** Investigation, Methodology, Resources, Writing – review & editing. **Michal Stanko:** Investigation, Writing – review & editing. **Bronislava Víchová:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing – review & editing. **Hein Sprong:** Conceptualization, Funding acquisition, Investigation, Supervision, Writing – review & editing.

Funding

The work of Bronislava Víchová, Veronika Blažeková and Michal Stanko was financially supported by the project VEGA 2/0051/24. Hein Sprong and Valérie Baede were financially supported by the Dutch Ministry of Health, Welfare and Sport, The Netherlands (VWS).

Declaration of competing interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ron P. Dirks, Hans J. Jansen, Wouter J. Veneman, and Jane Segobola are employed by the company Future Genomics Technologies B.V. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dana Zubriková, and Lucia Vargová for their help with field collections and the blood-feeding of ticks on mice in the laboratory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crpvbd.2026.100346.

Data availability

The mitochondrial genome and the raw read data were deposited at ENA and received the following accession numbers: Assembly_name: mtdna_v2.0; Assembly_acc: gca_965279885; Study_id: prjeb89026; Sample_id: ers24279335; Chromosome_acc: oz258157-oz258157.

References

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat. Methods 18, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

Donath, A., Jühling, F., Al-Arab, M., Bernhart, S.H., Reinhardt, F., Stadler, P.F., et al., 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 47, 10543–10552. https://doi.org/10.1093/nar/gkz833.

Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O., et al., 2021. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biol. 22, 120. https://doi.org/10.1186/s13059-021-02336-9.

Gahan, A.B., 1934. On the identities of chalcidoid tick parasites (Hymenoptera). Proc. Entomol. Soc. Wash. 36, 89–97.

Giannotta, M.M., Smith, I., Michie, M., Blasdell, K., Dunn, M., Nicholls, J., et al., 2024. Molecular characterisation of australasian *Ixodiphagus* (Hymenoptera; Encyrtidae; Encyrtinae) reveals unexpected diversity and a potential novel host switch. Int. J. Parasitol. 54, 743–753. https://doi.org/10.1016/j.ijpara.2024.09.001.

Hebert, P.D.N., Bock, D.G., Prosser, S.W.J., 2023. Interrogating 1000 insect genomes for NUMTs: a risk assessment for estimates of species richness. PLoS One 18, e0286620. https://doi.org/10.1371/journal.pone.0286620.

Johnson, J., Soehnlen, M., Blankenship, H.M., 2023. Long read genome assemblers struggle with small plasmids. Microb. Genom. 9, mgen001024. https://doi.org/10.1099/mgen.0.001024.

Kinkar, L., Gasser, R.B., Webster, B.L., Rollinson, D., Littlewood, D.T.J., Chang, B.C.H., et al., 2021. Nanopore sequencing resolves elusive long tandem-repeat regions in mitochondrial genomes. Int. J. Mol. Sci. 22, 1811. https://doi.org/10.3390/ijms22041811.

Krawczyk, A.I., Bakker, J.W., Koenraadt, C.J.M., Fonville, M., Takumi, K., Sprong, H., Demir, S., 2020. Tripartite interactions among *Ixodiphagus hookeri, Ixodes ricinus* and deer: differential interference with transmission cycles of tick-borne pathogens. Pathogens 9, 339. https://doi.org/10.3390/pathogens9050339.

Lin, Z.J., Wang, X., Wang, J., Tan, Y., Tang, X., Werren, J.H., et al., 2021. Comparative analysis reveals the expansion of mitochondrial DNA control region containing unusually high G-C tandem repeat arrays in *Nasonia vitripennis*. Int. J. Biol. Macromol. 166, 1246–1257. https://doi.org/10.1016/j.ijbiomac.2020.11.007.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W., O'Brien, S.J., 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. 39, 174–190. https://doi.org/10.1007/BF00163806.

Ramos, R.A.N., de Macedo, L.O., Bezerra-Santos, M.A., de Carvalho, G.A., Verocai, G.G., Otranto, D., 2023. The role of parasitoid wasps, *Ixodiphagus* spp. (Hymenoptera: Encyrtidae), in tick control. Pathogens 12, 676. https://doi.org/10.3390/pathogens12050676.

Samish, M., Ginsberg, H., Glazer, I., 2004. Biological control of ticks. Parasitology 129 (Suppl. l), S389–S403. https://doi.org/10.1017/S0031182004005219.

Xing, Z.-P., Liang, X., Wang, X., Hu, H.-Y., Huang, Y.-X., 2022. Novel gene rearrangement pattern in mitochondrial genome of *Ooencyrtus plautus* Huang & Noyes, 1994: new gene order in Encyrtidae (Hymenoptera, Chalcidoidea). ZooKeys 1124, 1–21. https://doi.org/10.3897/zookeys.1124.83811.

Xue, L., Moreira, J.D., Smith, K.K., Fetterman, J.L., 2023. The mighty NUMT: mitochondrial DNA flexing its code in the nuclear genome. Biomolecules 13, 753. https://doi.org/10.3390/biom13050753.