Journal of Species Research 15(2):198-206, 2026
https://doi.org/10.12651/JSR.2026.15.2.198

Open Access

The application of Nanopore long-read sequencing to reveal an exceptionally large tandem repeat region in the *Remanea naksanensis* mitochondrial genome

Jaehyun Kim[1] and Wonchoel Lee[2,*]

[1]*Advanced Infrastructure Operations Center, Korea Institute of Ocean Science & Technology, Busan 49111, Republic of Korea*
[2]*Department of Life Science, Hanyang University, Seoul 04763, Republic of Korea*

**Correspondent: wlee@hanyang.ac.kr*

Animal mitochondrial genomes are widely used as genetic markers in evolutionary studies such as systematics, species identification, and population genetics. Mitochondrial DNA is generally compact, encodes 37 genes, and typically contains a major non-coding region involved in replication and transcription. This region often harbors repetitive elements that hinder complete mitochondrial genome reconstruction using primer walking or short-read sequencing. In the present study, we first attempted to reconstruct the mitochondrial genome of *Remanea naksanensis* using Illumina short reads generated from three long-range PCR amplicons. However, the assemblies remained discontinuous in two putative non-coding regions. We subsequently employed Oxford Nanopore long-read sequencing in order to span these unresolved regions, and short reads were used for polishing. The resulting circular mitochondrial genome was found to be approximately 22.9 kb and contained two copies of an inverted repeat separated by a 9-bp spacer and an 8.2-kb tandem repeat array that could not be recovered from the short reads alone. This study provides an important mitochondrial genome resource for harpacticoid copepods, and a hybrid long- and short-read workflow for assembling repeat-rich mitochondrial genomes.

Keywords: Harpacticoida, long-read sequencing, mitogenome

Introduction

Mitochondrial DNA has been widely used as a molecular marker in animal evolutionary studies, including systematics, species identification, phylogeography, and population genetics (Avise, 1994; Dasalle *et al*., 2017). With a few exceptions, it is a closed circular molecule, approximately 15−20 kb in length. Many animal mitochondrial genomes contain a major non-coding region, generally termed the "control region", that includes regulatory elements for replication and transcription (Boore, 1999). This region often harbors repetitive sequences and segmental duplications (Gan *et al*., 2019; Formenti *et al*., 2021), which may contribute to mitogenome size variation (Savolainen *et al*., 2000; Wang *et al*., 2015) and can also hinder the assembly of complete mitochondrial genomes (Le *et al*., 2001; Gan *et al*., 2019).

A common traditional strategy for sequencing metazoan mitochondrial genomes involves long-range PCR coupled with primer walking (Dotson and Beard, 2001; Machida *et al*., 2002; Miller *et al*., 2004; Burger *et al*., 2007). Although primer walking is highly specific and yields

accurate sequences, it is relatively slow and expensive (Cameron, 2014). Recently, long-range PCR amplicons have been sequenced using next-generation (short-read) sequencing platforms rather than primer walking (Cameron, 2014; Gan *et al*., 2019). This time- and cost-efficient approach enables the high-coverage assembly of complete mitochondrial genomes (Shendure *et al*., 2017; Kinkar *et al*., 2020). However, short-read data often struggle to resolve the repetitive elements that exceed individual read lengths and lack unique flanking sequences on both sides (Sharbrough *et al*., 2023). Third-generation (long-read) sequencing (e.g., Oxford Nanopore technology) can generate reads longer than 10 kb without being affected by nucleotide composition (Rang *et al*. 2018; Gan *et al*., 2019). These extended reads can span several kilobases, revealing large repeat structures that cannot be assembled from short-read data (Shendure *et al*., 2017; Blom, 2021).

Copepoda is a small crustacean taxon that plays a pivotal role in aquatic ecosystems as a vital link between primary producers and higher trophic levels (Huys and Boxshall, 1991). With advances in sequencing technologies, the number of published copepod mitochondrial genomes has increased; however, the available data still represent only a small fraction of their extensive species diversity. In particular, although harpacticoids comprise about 30% of all known copepods (ca. 4,500 species; Walter and Boxshall, 2025) and are a key component of benthic communities (Wells, 1988; Boxshall and Halsey, 2004), complete harpacticoid mitochondrial genomes have been reported for only six species in three genera to date (*Amphiascoides atopus*, *Ikanecator primus*, *Tigriopus californicus*, *T. japonicus*, *T. kingsejongensis*, and *T. west*; Machida *et al*., 2002; Jung *et al*., 2006; Burton *et al*., 2007; Easton *et al*., 2014; Hwang *et al*., 2019; Zoral *et al*., 2024).

Herein, we present the complete mitochondrial genome of the harpacticoid copepod *Remanea naksanensis* Back, Lee & Huys, 2011. Initially, we attempted to assemble the mitochondrial genome using only the Illumina short reads generated from three long-range PCR amplicons; however, two putative non-coding regions remained unresolved. To overcome this limitation, we generated Oxford Nanopore long reads from the same long-range PCR products, enabling complete circularization and annotation of the mitogenome. We characterized two long non-coding regions, including two inverted-repeat copies separated by a 9-bp spacer, and an exceptionally large tandem repeat array with a terminal truncated repeat unit. Our study provides a validated mitochondrial genomic resource for harpacticoid copepods and describes the repeat-rich mitochondrial non-coding regions in harpacticoid copepods.

Materials and Methods

Sample collection and DNA extraction

A meiofaunal sample was collected on June 23, 2020, from a Karaman-Chappuis hole at the estuary of Munam Stream in Goseong, Korea (38°17′47″N, 128°32′56″E), with a fine-sand substrate. Paramesochrid individuals were isolated under a stereomicroscope (SApo, Leica microsystems, Wetzlar, Germany) and genomic DNA was extracted from each of several isolated individuals using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The remaining exoskeletons were retained for morphological species identification based on previous taxonomic studies (e.g., Huys *et al*., 1996; Wells, 2007).

Conventional amplification

Remanea naksanensis, the most abundant species in the meiofaunal sample, was selected as the target for mitochondrial genome amplification. Three mitochondrial gene fragments, cytochrome c oxidase subunit I (*COXI*), small subunit ribosomal RNA gene (*srRNA*), and cytochrome b (*COB*), were amplified from a single individual as anchor regions for subsequent long-PCR. Fragment amplification was performed using Premix EX Taq HS (Takara Bio, Shiga, Japan) with the primer pairs listed in Table 1. PCR products were sequenced using the same primers by a commercial Sanger sequencing service provider (Bionics, Seoul, Korea). All of the newly obtained sequences were assigned GenBank accession numbers (*COXI*: PX884362; *srRNA*: PZ024137; *COB*: PX892115).

Long-PCR and sequencing

Species-specific primer sets were designed to amplify the entire mitochondrial genome in three overlapping long-range PCR fragments based on anchor sequences (Table 1). Long-range PCR was performed in 25 μL reaction volumes using the same DNA template as in the anchor amplification, each primer pair, and LongAmp HS Taq Master Mix (New England Biolabs, Ipswich, MA, USA). The cycling conditions followed the manufacturer's recommendations, except that the extension step per cycle was set to 7−12 min depending on the estimated amplicon length, and the final extension step was 10 min. The PCR products were examined on a 1% agarose gel to confirm the expected amplicon size. The amplicons were sent to Bionics (Seoul, Korea) for paired-end sequencing on an Illumina MiSeq platform.

Despite the use of multiple strategies, including de novo assembly using SPAdes (v3.15.3; Prjibelski *et al*., 2020) and seed-based assembly using NOVOPlasty

Table 1. Information of the PCR primers used in this study.

Name	Sequences (5′-3′)	Tm (°C)	Reference
COXI			
jgLCO1490	TITCIACIAAYCAYAARGAYATTGG	54.6	Geller et al., 2013
cop-coi-2189	GGGTGACCAAAAAATCARAA	55.2	Bucklin et al., 2010
COB			
ucytb151F	TGTGGRGCNACYGTWATYACTAA	60.1	Merritt et al., 1998
ucytb270R	AANAGGAARTAYCAYTCNGGYTG	60.0	Merritt et al., 1998
srRNA			
L13337-12S	YCTACTWTGYTACGACTTATCTC	54.1	Machida et al., 2002
H13845-12S	GTGCCAGCAGCTGCGGTTA	64.7	Machida et al., 2002
COX1–COB			
sed115-01-01	AGCTTGATCAGGAATAATTGGCACAGGAC	64.3	This study
sed115-01-03	GAGGAGGGTTTGCTGTTGACAATGCTACT	66.1	This study
COB–srRNA			
sed115-01-04	CAGGAGTGACTAAAGGATTAGCTGGGATA	63.4	This study
sed115-01-06	GGGTTAGAAGCCAGGTATTACTGAAGTAA	61.5	This study
srRNA–COX1			
sed115-01-05	GAGGATTGTCTTCCTAACTCAACTTTCTC	61.3	This study
sed115-01-02	AAGAGAAGAATAGCCGTGACTAGGACAGA	64.1	This study

(v4.3.1; Dierckxsens *et al*., 2016), short-read assemblies remained discontinuous because of unresolved gaps within the two putative non-coding regions. In order to resolve these regions, additional long-range PCR amplifications were performed, and the amplicons were then sequenced on a MinION Mk1B (Oxford Nanopore Technologies, Oxford, UK). Libraries were prepared using the Ligation Sequencing Kit (SQK-LSK109) together with Native Barcoding Expansion 1-12 (EXP-NBD104). Equimolar amounts of barcoded libraries were pooled and loaded onto an R9.4.1 flowcell. Library preparation and sequencing steps were performed according to the manufacturer's instructions (NBA_9093_v109_revO_12NOV2019). Raw signals were basecalled with a super-accuracy model and demultiplexed using Guppy (v6.5.7; Oxford Nanopore Technologies, Oxford, UK).

Assembly and annotation

The read length distribution was initially assessed using NanoPlot (v1.46.2; de Coster and Rademakers, 2023) with 100-bp bins. The basecalled reads for each amplicon were filtered using fastplong (v0.4.1; Chen *et al*., 2018) to retain reads within the modal read length bin, and reads with more than 40% of bases having a Phred score ≤ 10 were discarded. The remaining reads were assembled de novo using Flye (v2.9.6; Kolmogorov *et al*., 2019) using the *--nano-hq* and *--asm-coverage 100* options and the estimated length of each amplicon. The minimum overlap length was set to approximately 30% of the estimated size. The assemblies of each amplicon were polished with four iterations of Racon (v1.5.0; Vaser *et al*., 2017) using the filtered reads, followed by Medaka (v2.1.1; Oxford Nanopore Technologies, Oxford, UK) using the *r941_min_sup_g507* model. Final polishing was performed with Pilon (v1.24, *--mindepth* 24; Walker *et al*., 2014) using Illumina short reads.

The three amplicon assemblies were manually stitched based on the overlapping regions guided by the anchor sequences. The complete mitochondrial sequence was annotated using the MITOS web server (v2.1.10; Bernt *et al*., 2013) and ARWEN (v1.2; Laslett and Canbäck, 2008), and protein-coding genes and ribosomal RNA gene boundaries were manually curated. Tandem and inverted repeats were identified using Tandem Repeat Finder (v4.0.9; Benson, 1999) and Inverted Repeat Finder (v3.09; Warburton *et al*., 2004), respectively. The mitochondrial genome of *R. naksanensis* was visualized using CIRCOS (v0.69-8; Krzywinski *et al*., 2009).

Results and Discussion

A total of 333,847 paired-end reads were generated on an Illumina MiSeq using three long-range PCR amplicons to assemble the mitochondrial genome of *R. naksanensis* (*COXI-COB*: 113,667 reads, SRR36883791; *COB-srRNA*:

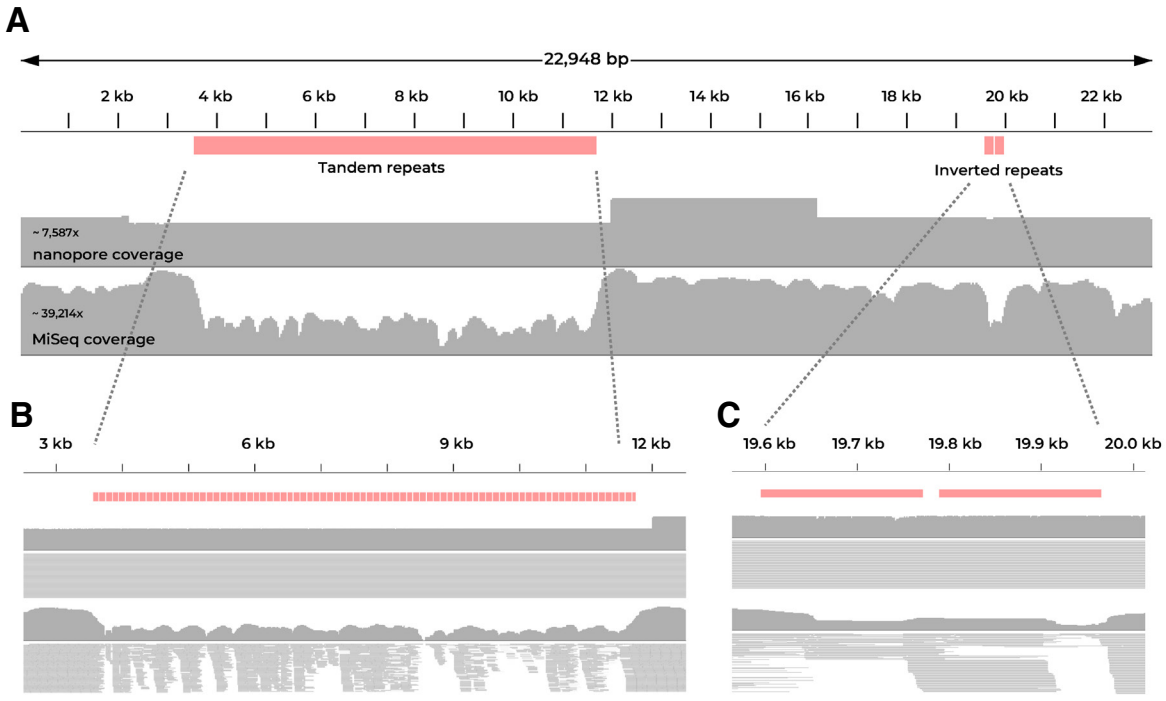

Fig. 1. IGV (Robinson et al., 2023) view showing the read-mapping evidence across repeat-rich non-coding regions of the *Remanea naksanensis* mitochondrial genome. A: Genome-wide overview of the read-depth profiles from Nanopore and MiSeq amplicon sequencing, highlighting the tandem repeat array and the inverted repeat (red blocks) together with coverage depth from Oxford Nanopore (top) and Illumina Miseq (bottom); B: Zoom-in of the tandem-repeat array showing the ambiguous short-read mapping across repetitive segments, whereas long reads support continuity across the array; C: Zoomed view of the inverted repeat region representing consistent long-read support across the repeat boundaries. Read alignments were cropped in IGV.

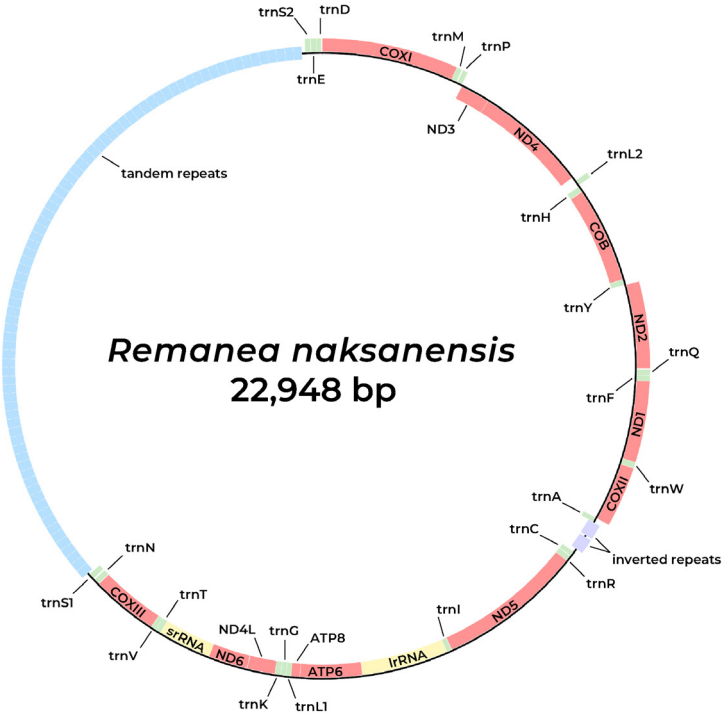

Fig. 2. Mitochondrial genome map of *Remanea naksanensis*. Genes encoded on the J- and N- strand are shown outer or inner circle, respectively.

105,883 reads, SRR36883787; *srRNA-COXI*: 114,297 reads, SRR36883789). Despite the high sequencing depth (>39,000×; Fig. 1A), the two regions remained discontinuous in both the seed-based and de novo assemblies. We inferred that the repetitive elements in these regions complicated assembly and were difficult to resolve using short-read data (Gan *et al*., 2019; Tørresen *et al*., 2019). Accordingly, we generated Oxford Nanopore long-read data from the same long-range PCR products in order to span the discontinuities. In total, 71,359 long reads were obtained (*COXI*-COB: 56,498 reads, SRR36883790; *COB-srRNA*: 8,318 reads, SRR36883786; *srRNA-COXI*: 6,543 reads, SRR36883788), yielding approximately 327.5 Mb of sequence data (*COXI-COB*: 223.6 Mb; *COB-srRNA*: 64.7 Mb; *srRNA-COXI*: 39.2 Mb). The overall read length N50 was 4,190 bp (*COXI-COB*: 4,169 bp; *COB-srRNA*: 9,217 bp; *srRNA-COXI*: 8,791 bp). Long-read data spanned both problematic regions (Fig. 1B, C), enabling the complete assembly of a single circular mitochondrial genome (Fig. 2).

The complete mitochondrial genome of *R. naksanensis* is 22,948 bp in length (accession number: PX893796; Fig. 2), with an A+T content of 76.2%. It contains a typical set of 37 mitochondrial genes (13 protein-coding genes, 22 tRNAs, and two rRNA genes). All protein-coding genes used ATN start codons and terminated with TAA stop codons, except for *COXI*, which ended with an incomplete stop codon (T) immediately preceding *trnM* (Table 2). The J-strand encoded 24 genes, whereas the N-strand encoded 13 genes. In addition to this canonical architecture, two non-coding regions longer than 100 bp were identified (Figs. 1, 2; Table 2), which together accounted for a substantial proportion of the increased genome size relative to other harpacticoid mitochondrial genomes. At the boundary of each non-coding region, a repetitive unit overlaps the adjacent tRNA by a few nucleotides.

One non-coding region located between *trnA* and *trnC* harbored two inverted repeat copies (Fig. 2; Table 2). This region was A+T rich (approximately 80%). The two copies were 181 bp each, separated by a 9-bp spacer, and showed approximately 99% sequence identity with two mismatches and no indels. The other non-coding region, located between *trnS1* and *trnS2*, contained a long tandem repeat array composed of approximately 100-bp repeat units (A+T content = ~90%). Long-read sequencing enabled the characterization of this region, identifying 81 repeat units (97–105 bp) and a terminally truncated repeat unit (41 bp) adjacent to *trnS2*. Overall, the tandem repeat array exceeded 8 kb, representing, to our knowledge, the longest single non-coding region reported in copepod mitochondrial genomes to date (the former longest: 4,341 bp in *Eurytemora affinis*, NC_046694; Choi *et al*., 2019).

The alignment of the repeat units revealed five conserved sequence blocks with only minor variations (1–3 mismatches) across the otherwise conserved sites (Fig. 3). The 41-bp truncated repeat unit contained the first three conserved blocks but terminated prematurely relative to the full-length units. The presence of this truncated unit at the array boundary suggests that it was generated via replication slippage. Such boundary-imperfect units commonly attributed to slipped-strand mispairing, a mecha-

Table 2. Mitochondrial genome organization of *Remanea naksanensis*.

Name	Location	Strand	Size	Codon		Intergenic nucleotides
				Start	Stop	
COXI	1-1546	N	1,546	ATC	T	0
trnM	1547-1608	N	62			10
trnP	1619-1682	N	60			6
ND3	1689-2039	J	351	ATA	TAA	−1
ND4	2039-3379	J	1,341	ATA	TAA	50
trnL2	3430-3496	N	67			11
trnH	3508-3578	J	71			2
COB	3581-4714	J	1,134	ATG	TAA	1
trnY	4716-4782	J	67			1
ND2	4784-5767	N	984	ATT	TAA	11
trnQ	5779-5846	N	68			1
trnF	5848-5909	N	62			−3
ND1	5907-6833	N	927	ATA	TAA	−1
trnW	6833-6899	N	67			1
COXII	6901-7596	N	696	ATG	TAA	−5
trnA	7592-7653	J	62			417
trnC	8069-8127	J	59			5
trnR	8133-8196	J	64			1
ND5	8198-9913	J	1,716	ATA	TAA	2
trnI	9916-9977	J	62			0
lrRNA	9978-11005	J	1,028			0
ATP6	11006-11746	J	741	ATG	TAA	−8
ATP8	11739-11843	J	105	ATA	TAA	−1
trnL1	11843-11904	J	62			−1
trnG	11904-11964	J	61			2
trnK	11967-12034	J	68			0
ND4L	12035-12364	J	330	ATT	TAA	3
ND6	12368-12832	J	465	ATA	TAA	0
srRNA	12833-13466	J	634			0
trnT	13467-13530	J	64			0
trnV	13531-13592	J	62			0
COXIII	13593-14384	J	792	ATG	TAA	5
trnN	14390-14458	J	69			18
trnS1	14477-14542	J	66			8205
trnS2	22748-22813	N	66			3
trnE	22817-22881	N	65			0
trnD	22882-22944	N	63			4

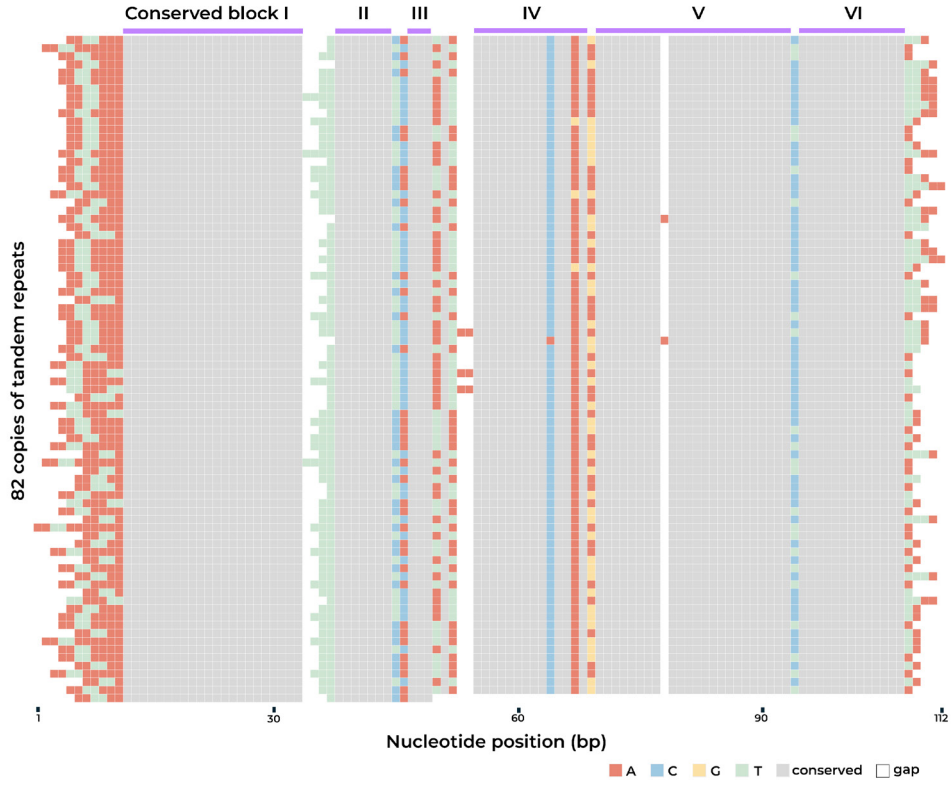

Fig. 3. Sequence conservation across the 82 tandem repeat copies extracted from the expanded non-coding region.

nism widely invoked for imprecise or pre-mature termination of replication, recombination during replication, and/or maintenance (Levinson and Gutman, 1987; Lunt and Hyman, 1997; Pâques *et al*., 1998; Rokas *et al*., 2003).

In the non-coding regions of animal mitochondrial genomes, sequence and length variations are commonly observed within species, and length heteroplasmy has been widely reported (Rand, 1993; Lunt *et al*., 1998). Such variations in length have important implications for evolutionary studies (Zhang and Hewitt, 1997). Although similar variability may occur in the two non-coding regions of *R. naksanensis*, our data were generated from long-range amplicons derived from a single individual, which precluded direct assessment of intraspecific variation. The read-mapping results did not reveal any obvious evidence of repeated copy number heterogeneity in this individual (Fig. 1). However, additional sequencing without PCR amplification and read-length binning is required to evaluate repeat-length polymorphisms and repeat-unit diversity at both the individual and population levels.

In the present study, we have assembled the complete mitochondrial genome of *R. naksanensis* de novo using a combination of short (Illumina MiSeq) and long (Oxford Nanopore sequencing) reads generated from three long-range PCR amplicons. This mitochondrial genome provides a validated resource for further comparative and phylogenetic studies on harpacticoid copepods, including paramesochid lineages, for which mitochondrial genomic data remain limited. Furthermore, our long-read approach revealed an approximately 8-kb tandem repeat region that was initially undetected in short-read assemblies.

Most of the reported copepod mitochondrial genomes have been sequenced using Sanger sequencing or short-read platforms, and some remain only partially characterized (e.g. *Calanus sinicus* (GU355641), Minxiao *et al*., 2011; *Calanus fimarchicus* (MG001887) and *Calanus glacialis* (MG001883), Weydmann *et al*., 2017; *Schizopera knabeni* (KF667527), Easton *et al*., 2014). Recently, concerns have been raised that mitochondrial genomes assembled solely from short, even when reported as complete, may omit repeat-rich regions and thus remain in complete (Filipović *et al*., 2021; Formenti *et al*., 2021; Kinkar *et al*., 2021). Consistent with this, several complete mitochondrial genomes based on short reads have later been revised using long-read sequencing, which enabled the detection of previously unrecognized structures within non-coding regions (Kinkar *et al*., 2020; Sharbrough *et al*., 2023). Therefore, long-read sequencing is an important strategy for accurately characterizing repeat-rich mitochondrial genomes and improving assembly completeness.

Acknowledgements

This research was funded by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, grant number 2021R1I1A2043807 and includes results obtained using research instruments operated by the Marine Core Facility at the Korea Institute of Ocean Science and Technology.

Conflicts of Interest

The author of this paper have no affiliation with any interests and are solely responsible for the paper.

References

Avise, J. 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York, xiv+511pp.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acid Research 27(2):573-580.

Bernt, M., A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf and P.F. Stadler. 2013. MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation. Molecular Phylogenetics and Evolution 69(2):313-319.

Blom, M.P.K. 2021. Opportunities and challenges for high-quality biodiversity tissue archives in the age of long-read sequencing. Molecular Ecology 30:5935-5948.

Boore, J.L. 1999. Animal mitochondrial genomes. Nucleic Acids Research 27(8):1767-1780.

Boxshall, G.A. and S.H. Halsey. 2004. An introduction to copepod diversity. The Ray Society, London. 966pp.

Bucklin, A., B.D. Ortman, R.M. Jennings, L.M. Nigro, C.J. Sweetman, N.J. Copley, T. Sutton and P.H. Wiebe. 2010. A "Rosetta Stone" for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). Deep Sea Research Part II: Topical Studies in Oceanography 57(24-26):2234-2247.

Burger, G., D. Lavrov, L. Forget and F. Lang. 2007. Sequencing complete mitochondrial and plastid genomes. Nature Protocols 2(3):603-614.

Burton, R.S., R.J. Byrne and P.D. Rawson. 2007. Three divergent mitochondrial genomes from California populations of the copepod *Tigriopus californicus*. Gene 403 (1-2):53-49.

Cameron, S.L. 2014. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Systematic Entomology 39:400-411.

Chen, S., Y. Zhou, Y. Chen and J. Gu. 2018. Fastp: and ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17):i884-i890.

Choi, B.S., J. Han, D.S. Hwang, S. Souissi, A. Hagiwara and J.S. Lee. 2019. Complete mitochondrial genome of the calanoid copepod *Eurytemora affinis* (Calanoida, Temoridae). Mitochondrial DNA Part B. 4(2):2731-2733.

DaSalle, R., B. Schierwater and H. Hadrys. 2017. MtDNA: The small workhorse of evolutionary studies. Frontiers in Bioscience (Landmark Ed) 22:873-887.

de Coster, W.D. and R. Rademakers. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39(5):btad311.

Dierckxsens, N., P. Mardulyn and G. Smits. 2016 NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45(4):e18.

Dotson, E.M. and C.B. Beard. 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, Triatoma dimidiate. Insect Molecular Biology 10:205-215.

Easton, E.E., E.M. Darrow, T. Spears and D. Thistle. 2014 The mitochondrial genomes of *Amphiascoides atopus* and *Schizopera knabeni* (Harpacticoida: Miraciidae) reveal similarities between the copepod orders Harpacticoida and Poecilostomatoida. Gene 538(1):123-137.

Filipović, I., J.P. Hereward, G. Rašić, G.J. Devine, M.J. Furlong and K. Etebari. 2021. The complete mitochondrial genome sequence of *Oryctes rhinoceros* (Coleoptera: Scarabaeidae) based on longread nanopore sequencing. PeerJ 9:e10552.

Formenti, G., A. Rhie, J. Balacco, B. Haase, J. Mountcastle, O. Fedrigo, S. Brown, M.R. Capodiferro, F.O. Al-Ajli, R. Ambrosini, P. Houde, S. Koren, K. Oliver, M. Smith, J. Skelton, E. Betteridge, J. Dolucan, C. Corton, … E.D. Jarvis. 2021. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biology 22(1):120.

Gan, H.M., S.M. Linton and C.M. Austin. 2019. Two reads to rule them all: Nanopore long read-guided assembly of the iconic Christmas Island red crab, *Gecarcoidea natalis* (Pocock, 1888), mitochondrial genome and the challenges of AT-rich mitogenomes. Marine Genomics 45:64-71.

Geller, J., C. Meyer, M. Parker and H. Hawk. 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular Ecology Resources 13(5):851-861.

Huys, R. and G.A. Boxshall. 1991. Copepod Evolution. The Ray Society, London. 468pp.

Huys, R., J.M. Gee, C.G. Moore and R. Hamond. 1996. Marine and brackish water harpacticoid copepods. Part 1. Keys and notes for identification of the species. In: R.S.K. Barnes and J.H. Crothers (eds.). Synopses of the British Fauna, New Series, Published for the Linnean Society of London and the Estuarine and Coastal Sciences Association 51(1):1-352.

Hwang, D.S., B.S. Choi, M.C. Lee, J. Han, S. Kim and J.S. Lee. 2019. Complete mitochondrial genome of the Antarctic copepod *Tigriopus kingsejongenesis* (Harpacticoida, Harpacticidae). Mitochondrial DNA Part B 4(1):1470-1471.

Jung S.-O., Y.-M. Lee, T.-J. Park, H.G. Park, A. Hagiwara, K.M.Y. Leung, H.-U. Dahms, W. Lee and J.-S. Lee. 2006. The complete mitochondrial genome of the intertidal copepod *Tigriopus* sp. (Copepoda, Harpactidae) from Korea and phylogenetic considerations. Journal of Experimental Marine Biology and Ecology 333:251-262.

Kinkar, L., N.D. Young, W.-M. Sohn, A.J. Stroehlein, P.K. Korhonen, R.B. and Gasser. 2020. First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. PLOS Neglected Tropical Diseases 14(8):e0008552.

Kinkar, L., R.B. Gasser, B.L. Webster, D. Rollinson, D.T.J. Littlewood, B.C.H. Chang, A.J. Stroehlein, P.K. Korhonen and N.D. Young. 2021. Nanopore Sequencing Resolves Elusive Long Tandem-Repeat Regions in Mitochondrial Genomes. International Journal of Molecular Sciences 22:1811.

Kolmogorov, M., J. Yuan, Y. Lin and P.A. Pevzner. 2019. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology 37:540-546.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones and M.A. Marra. 2009. Circos: an information aesthetic for comparative genomics. Genome Research 19(9):1639-1645.

Laslett, D. and B. Canbäck, B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24(2):172-175.

Le T.H., P.-F. Humair, D. Blair, T. Agatsuma, D.T.J. Littlewood and D.P. McManus. 2001. Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. Molecular and Biochemical Parasitology 117:61-71.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution 4(3):203-221.

Lunt, D.H. and B.C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature 387:247.

Lunt, D.H., L.E. Whipple and B.C. Hyman. 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Molecular Ecology 7:1441-1455.

Machida, R.J., M.U. Miya, M. Nishida and S. Nishida. 2002. Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). Marine Biotechnology 4:406-417.

Merritt, T.J., L. Shi, M.C. Chase, M.A. Rex, R.J. Etter and J.M. Quattro. 1998. Universal cytochrome b primers facilitate intraspecific studies in molluscan taxa. Molecular Marine Biology and Biotechnology 7(1):7-11.

Miller, A.D., T.T. Nguyen, C.P. Burridge and C.M. Austin, 2004. Complete mitochondrial DNA sequence of the Australian freshwater crayfish, *Cherax destructor* (Crustacea: Decapoda: Parastacidae): a novel gene order revealed. Gene 331:65-72.

Minxiao, W., S. Song, L. Chaolun and S. Xin. 2011. Distinctive mitochondrial genome of Calanoid copepod *Calanus sinicus* with multiple large non-coding regions and reshuffled gene order: Useful molecular markers for phylogenetic and population studies. BMC Genomics 12:73.

Pâques, F., W.-Y. Leung and J.E. Haber. 1998. Expansions and contractions in a tandem repeat induced by double-strand break repair. Molecular and Cellular Biology 18:2045-2054.

Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus and A. Korobeynikov. 2020. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70(1):1-29.

Rand, D.M. 1993. Endotherms, ectotherms, and mitochondrial genome-size variation. Journal of Molecular Evolution 37(3):281-295.

Rang, F.J., W.P. Kloosterman and J. de Ridder. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biology 19:90.

Robinson, J.T., H. Thorvaldsdóttir, D. Turner and J.P. Mesirov. 2023. igv.js: an embeddable JavaScript implementation of the Integrative Genomics View (IGV). Bioinformatics 39(1):btac830.

Rokas, A., E. Ladoukakis and E. Zouros. 2003. Animal mitochondrial DNA recombination revisited. Trends in Ecology & Evolution 18(8):411-417.

Savolainen, P., L. Arvestad and L. Lundeber. 2000. mtDNA Tandem Repeats in Domestic Dogs and Wolves: Mutation Mechanism Studied by Anlaysis of the Sequence of Imperfect Repeats. Molecular Biology and Evolution 17(4):474-488.

Sharbrough, J., L. Banker, E. Cook, P.D. Fields, J. Jalinsky, K.E. McElroy, M. Neiman, J.M. Logsdonjr and J.L. Boore. 2023. Single-molecule Sequencing of an Animal Mitochondrial Genome Reveals Chloroplast-like Architecture and Repeat-mediated Recombination. Molecular Biology and Evolution 40(1):msad007.

Shendure, J., S. Balasubramanian, G.M. Church, W. Gilbert, J. Rogers, J.A. Schloss and R.H. Waterston. 2017. DNA sequencing at 40: past, present and future. Nature 2017: 550(7676):345-353.

Tørresen, O.K., B. Star, P. Mier, M.A. Andrade-Navarro, A. Bateman, P. Jarnot, A. Gruca, M. Crynber, A.V. Kajava, V.J. Promponas, M. Anisimova, K.S. Jakobsen and D. Linke. 2019. Tandem repeats lead to sequence assembly errors and impose multi-level challenge for genome and protein databases. Nucleic Acids Research 47921:10994-11006.

Vaser, R., I. Sović, N. Nagarajan and M. Šikić. 2017. Fast and accurate de novo genome assembly from long uncorrected

reads. Genome Research 27(5):737-746.

Walker, B.J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young and A.M. Earl. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE 9(11):e112963.

Walter, T.C. and G. Boxshall. 2025. World of Copepods Database. [Available from: https://www.marinespecies.org/copepoda/ on 2025-11-16].

Wang, X., N. Liu, H. Zhang, X.-J. Yang, Y. Huang and F. Lei. 2015. Extreme variation in patterns of tandem repeats in mitochondrial control region of yellow-browed tits (*Sylviparus modestus*, Paridae). Scientific Reports 5:13227.

Warburton, P.E., J. Giordano, F. Cheung, Y. Gelfand and G. Benson. 2004. Inverted Repeat Structure of the Human Genome: The X-Chromosome Contains a Preponderance of Large, Highly Homologous Inverted Repeats That Contain Testes Genes. Genome Research 14:1861-1869.

Wells, J.B.J. 1988. Copepoda. In: R.P. Higgins and H. Thiel (eds.). Introduction to the study of meiofauna. Smithsonian Institution Press, Washington D.C., 380-388.

Wells, J.B.J. 2007. An annotated checklist and keys to the species of Copepoda Harpacticoida (Crustacea). Zootaxa 1568:1-872.

Weydmann, A., A. Przyłucka, M. Lubośny, K.S. Walczyńska, E.A. Serrão, G.A. Pearson and A. Burzyński. 2017. Mitochondrial genomes of the key zooplankton copepods Arctic *Calanus glacialis* and North Atlantic *Calanus finmarchicus* with the longest crustacean non-coding regions. Scientific Reports 7:13702.

Zhang, D.-X. and G.M. Hewitt. 1997. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics and Ecology 2592:99-120.

Zoral, M.A., Z. Lajbner, L. Zifcakova and J. Miller. 2024. Peracetic acid treatment of squid eggs infected with parasitic copepod (*Ikanecator primus* gen. et sp. nov.). Scientific Reports 14:14513.

Submitted: February 2, 2026
Revised: February 19, 2026
Accepted: February 26, 2026