

RESOURCE ARTICLE OPEN ACCESS

Fragmentation of Long Reads Enables Reliable Mitogenome Assembly From Whole-Genome Amplification Data With Pervasive Palindromic Reads

Matteo Vecchi^{1,2}  | Bartłomiej Surmacz²  | Ingemar K. Jönsson³  | Daniel Stec² 

¹Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy | ²Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Kraków, Poland | ³Department of Environmental Science, Kristianstad University, Kristianstad, Sweden

Correspondence: Matteo Vecchi (matteo.vecchi@unipr.it; matteo.vecchi15@gmail.com)

Received: 26 March 2026 | **Revised:** 14 May 2026 | **Accepted:** 2 June 2026

Keywords: Acutuncus | chimeric sequences | long reads | mitochondria | nanopore | tardigrades

ABSTRACT

Whole genome amplification (WGA), and in particular multiple displacement amplification (MDA), has become a key technique for genomic sequencing of microscopic organisms, yet it introduces artefacts such as palindromic (inverted chimeric) reads that may compromise downstream analyses. We assessed how pervasive palindromic reads generated by MDA impact the assembly of tardigrade (*Acutuncus giovanninae* and *A. mecnuffi*) mitogenomes sequenced with Oxford Nanopore technology. We show that the MDA produces a high proportion of palindromic reads, often exceeding one-third of mitochondrial reads and frequently exhibiting complex multi-inversion structures. These artefacts severely impair long-read assembly, leading to low success rates and inconsistent genome reconstruction. To solve this issue, a strategy based on in silico fragmentation of long reads into short, high-quality fragments, followed by short-read assembly, consistently produced complete and accurate circularised mitochondrial genomes. Our results demonstrate that palindromic read formation can be, in some cases, a limitation of MDA coupled with long-read sequencing, but this issue can be mitigated through read fragmentation. This approach provides a simple, robust and scalable solution for mitogenome assembly from data heavily affected by amplification artefacts, particularly in microscopic taxa where whole genome amplification is often unavoidable.

1 | Introduction

The use of mitochondrial genomes in resolving phylogenetic relationships among different animal taxa has undoubtedly brought progress to the field of phylogenetics and systematics (e.g., Forni et al. 2021; García-Souto et al. 2025; Lavrov et al. 2023; Sun et al. 2020). While first restricted by the inherent limitations of Sanger sequencing (Zardoya and Suárez 2008), the advent of second (454, Illumina and IonTorrent sequencing) and third (PacBio and Oxford Nanopore) generation massive parallel sequencing technologies were the principal drivers behind the sequencing and release of a massive amount of new genomic data. Among the third-generation sequencing technologies, the small and cheap Oxford Nanopore Technologies MinION sequencer allowed even

small laboratories to obtain high-quality genomic data (Menegon et al. 2017; Parker et al. 2017). Yet a major limitation of these sequencing technologies is the amount of high-molecular-weight input DNA required (Pollard et al. 2018). While, in principle, enough material for library preparation and sequencing could be obtained from a single fly (Kim et al. 2024), obtaining enough DNA for microscopic metazoans or single-cellular organisms is usually not possible. Although pooling of individuals is a potential solution (Flot et al. 2013; Li et al. 2024), it can heavily reduce the assembled genome quality due to high heterozygosity and contaminations (Bemm et al. 2016; Huang et al. 2017; Simion et al. 2018).

Whole genome amplification (WGA) refers to a series of methodologies that enable non-selective amplification of the whole genome

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

sequence (Wang et al. 2022). In particular, multiple displacement amplification (MDA) relies on isothermal amplification with Phi29 DNA polymerase and random exonuclease-resistant primers (Dean et al. 2002), enabling high yield and accuracy from a small amount of the starting DNA template, combined with a straight-forward laboratory protocol (Wang et al. 2022). However, this methodology has known limitations, including high variance in amplification efficiency (high variance of coverage and sequencing depth) across genomic regions due to GC amplification biases and differential binding efficiency of primers (Benita 2003; Borgström et al. 2017; Lee et al. 2023). Another well-recognised issue with MDA is the production of inverted chimeric sequences consisting of one or more inverted repeats (palindromes) when 3' termini are displaced and annealed to nearby displaced 5' strands (Lasken and Stockwell 2007; Lu et al. 2023; Sabina and Leamon 2015; Warris et al. 2018). Bioinformatic pipelines have been developed to find and correct palindromic reads (Agyabeng-Dadzie et al. 2025; Lee et al. 2023; Lu et al. 2023; Warris et al. 2018). However, one of these methods relies on mapping the reads to a reference genome (Lu et al. 2023), which limits its utility to already well-characterised organisms. Other methods only identify palindromes, which are then discarded completely (Agyabeng-Dadzie et al. 2025). Some approaches aim to correct palindromic reads (Lee et al. 2023) by splitting them in half in each software run iteration. Importantly, this approach is limited in its efficacy to situations where reads experienced only one or two palindrome formation events.

The development and commercialisation of WGA kits was particularly welcomed in the field of meiofauna research. The small size of meiofauna impacts both the effort put into their research due to difficulties in finding and identifying them (Fontaneto et al. 2015; Mammola et al. 2023; Martínez et al. 2025) and the amount of DNA available for massive parallel sequencing, leading to their underrepresentation in genomic and mitogenomic sequencing projects and studies (Kocot et al. 2017; Laumer et al. 2019). It comes naturally that there is a rise in the adoption of WGA for the sequencing of meiofauna genomes (e.g., Kosakyan et al. 2026; Roberts et al. 2024; Vecchi and Stec 2025). Therefore, the commercial availability of MDA kits brought forward the sequencing

of genomes and mitogenomes of meiofauna from single individuals, with both second- and third-generation massive parallel sequencing technologies (Camarda et al. 2025; Gammuto et al. 2024; Kosakyan et al. 2026; Lee et al. 2023; Roberts et al. 2024; Vecchi and Stec 2025). However, an investigation into how palindromic reads affect mitogenome assembly has not been performed.

Given the possibility to scale up the sequencing of mitogenomes of meiofauna provided by MDA, we used tardigrades as a focal system because they are microscopic animals for which single-individual mitogenome sequencing is often constrained by low DNA input and therefore depends on WGA-based approaches. We investigate how MDA with a high reaction gain coupled with long-read sequencing (Oxford Nanopore) is impacted by the formation of palindromic reads, and we propose an approach, based on splitting long reads into short fragments that mimic the output of short-read sequencing technologies, to efficiently produce mitogenome assemblies from data heavily impacted by this issue.

2 | Material and Methods

2.1 | Samples

Live tardigrades belonging to the species *Acutuncus giovanninae* Vecchi et al. 2023 (three individuals) and *Acutuncus mecnuffi* Vecchi et al. 2023 (two individuals) were obtained from rock pool sediments collected during a previous study (Vecchi et al. 2026). Sample information is provided in Table 1. Tardigrades were isolated and manipulated with flame-sterilised tungsten Irwin loops and kept overnight in sterile distilled water at 8°C to empty their gut contents.

2.2 | Whole Genome Amplification and Sequencing

Tardigrades were transferred with a sterilised tungsten Irwin loop into 4 µL of REPLI-g Advanced Storage buffer on a sterilised

TABLE 1 | Information overview of the sequenced individuals and sequencing run statistics, divided by individual and statistics for reads identified, as of putative mitochondrial origin by DIAMOND blastp.

Sample	Species	Country	Coordinates	Sum Gb raw reads [N50 bp—average quality Q]	Sum Gb mitochondrial reads [N50 bp—average quality Q]	% mitochondrial reads
Acu_gio_RIT094	<i>Acutuncus giovanninae</i>	Italy	44.381019, 10.047973	18.11 [5509–19.20]	0.16 [8126–17.80]	0.73
Acu_gio_RPL027	<i>Acutuncus giovanninae</i>	Poland	50.777043, 15.590996	16.99 [6637–19.46]	2.34 [5850–19.68]	15.78
Acu_gio_RSE010	<i>Acutuncus giovanninae</i>	Sweden	56.138949, 14.831353	13.18 [5041–19.37]	6.01 [4370–19.08]	52.26
Acu_mec_RSE086	<i>Acutuncus mecnuffi</i>	Sweden	56.155758, 14.876410	13.02 [7123–19.39]	0.51 [6998–19.00]	4.81
Acu_mec_RSE087	<i>Acutuncus mecnuffi</i>	Sweden	56.156057, 14.877798	12.70 [7612–19.63]	0.16 [8997–18.53]	1.38

glass slide and cut with an insulin needle to release the body cavity content, and then, the buffer containing the cells was transferred to a microcentrifuge vial. Whole genome amplification was then done following the REPLI-g Advanced DNA Single Cell Kit manufacturer's protocol. The reaction product was purified with a GeneMAGNET PCR/DNA Clean-Up Purification Kit (EURx), and the ds-DNA was quantified using a Qubit Fluorometric assay. Approximately 2 µg of amplified DNA was debranched for 1 h with T7 Endonuclease I (New England Biolabs) according to the manufacturer's protocol. The reaction product was again purified with GeneMAGNET and quantified with Qubit. About 1000 ng of the amplified DNA for each sample was used to prepare a library according to the Native Barcoding Kit 24 V14 SQKNBD114.24 protocol (Oxford Nanopore Technologies Limited, Oxford, UK). All samples were processed fully independently until after the addition of barcodes. The samples were combined into two pools, loaded on R10.4.1 PromethION flow cells (FLO-PRO114) and sequenced on a PromethION 24 instrument (MinKnow 24.06.15). Basecalling was performed using the dorado basecaller server 7.4.14 set to 'super-accurate' mode, 400 bp, 5 Hz. Reads with $Q < 10$ were discarded. Library preparation, sequencing and demultiplexing were performed by a commercial provider (IGA Technology, Udine, Italy). Residual adapters were removed with Porechop_ABI (Bonenfant et al. 2023).

2.3 | Assembly

To speed up the assembly process, mitochondrial reads were first extracted as in Vecchi and Stec (2025): in brief, the reads were blasted with DIAMOND blastx (Buchfink et al. 2021) against a database of tardigrade mitochondrial proteins (see Vecchi and Stec 2025) and reads with at least one match were retained. Two assembly approaches were then tested: one with long reads (hereafter referred to as LR assembly), and one based on fragmentation of long reads into short reads (hereafter referred to as SR assembly).

2.4 | LR Assemblies

Multiple assemblies per sample were produced with the software Flye v 2.9.3 (Kolmogorov et al. 2020; Lin et al. 2016). The first assembly was produced with the default Flye mode with 1 polishing iteration. The second was produced with the Flye—meta option with 1 polishing iteration, and the third was produced with the Flye—gsize option (genome size specified to 14.5 kb based on preliminary assembly tests) with 1 polishing iteration. The assembly process, as explained, was then repeated for the unmodified reads and for the reads with palindromes removed with the script provided by Lee et al. (2023), with two different minimum alignment lengths to classify a read as palindromic of 100 and 1000 bp. The second length of 1000 bp is the default value proposed in the original script of Lee et al. (2023), while 100 bp was chosen to provide a much more extreme palindrome detection. The process generated a total of nine assemblies for each sample in a crossed design: assembly algorithm (Flye standard, Flye meta, Flye gsize) * palindrome treatment (no treatment, alignment length 1000, alignment length 100). Assembly success was evaluated by the presence of circular contigs with a size close to 14.5 kb. The identity of the contigs (i.e., the contig

is a mitochondrial genome and belongs to the right species) was then confirmed by blasting the sequence against the NCBI core_nt database (Sayers et al. 2025) with blastn (Camacho et al. 2009) using default parameters.

2.5 | SR Assembly

To use an assembler designed for short-read sequencing technology (i.e., Illumina) before assembly, the reads were chopped into 150 bp fragments, and only fragments $Q \geq 30$ were retained using seqkit v 2.10.1 (Shen et al. 2016). Then, the mitogenomes were assembled with NOVOPlasty v.4.3.5 with k-mer size 33 (Dierckxsens et al. 2016) using COI sequences from either *A. giovanninae* or *A. mecnuffi* from GenBank (MW306855 Vecchi et al. 2022; OM350039 Vecchi et al. 2023). Assembly success was evaluated as above.

2.6 | Mitogenomes Annotation

The mitochondrial genomes were annotated with MITOS2 (Bernt et al. 2013; Donath et al. 2019). The annotation was inspected and curated by hand, and a GenBank flat file was produced using the scripts mitos2fasta.py and aln2tbl.py (<https://github.com/IMEDEA/mitogenomics>) and table2asn (https://ftp.ncbi.nlm.nih.gov/asn1-converters/by_program/table2asn/). The mitogenomes were visualised with OGDRAW (Greiner et al. 2019).

2.7 | Palindromes Quantification in Reads

Palindromes were quantified by blasting the long reads against the respective mitogenome assembly with blastn (Camacho et al. 2009; E-value threshold $1e-5$, identity threshold 95%). A read was considered palindromic if multiple mitogenome matches were present in different orientations. Processing of blastn results and results visualisation were performed with 'R v 4.5.2' (R Core Team 2022) using the packages 'tidyverse v 2.0.0' (Wickham et al. 2019), 'ggplot2 v 4.0.2' (Wickham 2016), 'ggridges v 0.5.7' and 'gggenes v 0.6.0'.

2.8 | Code Availability

All the code used for this study is available at <https://doi.org/10.5281/zenodo.19131525> and as Supporting Information SM.01.

3 | Results

The sequencing run produced about 12–18 Gb of sequences with average quality above Q19 (see Table 2). Read length N50 was between 3743 and 5304 bp, which can be considered low for standard Nanopore whole genome sequencing runs. This was expected as the fragmentation of the DNA library was worsened by the WGA and debranching process. The raw reads were submitted to NCBI SRA (Bioproject PRJNA1287536).

The number of putative mitochondrial reads identified by blastx against a tardigrade mitochondrial protein database varied greatly

TABLE 2 | Mitogenome assembly statistics for the different methods expressed as assembly size in bp [coverage].

	Flye (LR)									
	Raw reads			Palindrome 1000			Palindrome 100			
	Flye	Flye meta	Flye gsize	Flye	Flye meta	Flye gsize	Flye	Flye meta	Flye gsize	NOVOplasty
Acu_gio_RSE010	n.o.	14,913 [38]	n.o.	n.o.	14,398 [63]	n.o.	n.o.	n.o.	n.o.	14,416 [343,268]
Acu_gio_RPL027	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	14,417 [128,066]
Acu_gio_RIT094	n.o.	n.o.	n.o.	14,469 [663]	14,462 [682]	n.o.	14,470 [810]	14,440 [635]	n.o.	14,415 [4113]
Acu_mec_RSE086	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	14,398 [22,115]
Acu_mec_RSE087	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	14,399 [4325]

Abbreviation: n.o., indicates that a circular mitogenome of the approximately right size and of the right species was not obtained.

across samples (Table 1), from 0.73% to 52.23% of all reads. The proportion of mitochondrial reads measured as the number of reads with a blast match against the final mitogenome assembly (see paragraphs below) produced similar numbers identified with the method mentioned above (Table 1).

3.1 | LR Assembly

Circularised assemblies were produced by Flye only for two individuals (Table 2) belonging to *A. giovanninia*. Across the same individual, non-negligible differences in assembly size are found (ranging from 14,398 to 14,913 bp in *A. giovanninia*, surpassing any plausible intraspecific variability). Flye in genome size mode never produced an assembly, while Flye in metagenome mode performed the best. Due to the low success rate, heterogeneity in assembly lengths and relatively low coverage compared to the SR assemblies (see below), the obtained LR assemblies were considered unsuccessful and not used for other downstream analyses.

3.2 | SR Assembly and Mitogenome Annotation

The short-reads assembly method produced a circularised assembly of the expected size for each individual (Table 2), with negligible length differences between individuals of the same species (3 bp for *A. giovanninia* and 1 bp for *A. mecnuffi*). The automatic annotation with MITOS2 of the mitogenomes obtained with the SR assembly produced almost complete annotations (13 protein coding genes, 22 tRNAs, two ribosomal RNAs; Figure 1), with the exceptions of the absence of ATP synthase subunit 8 (atp8) in *A. giovanninia* (which was found to be present by a manual curation) and the false positive presence of two copies of NADH dehydrogenase subunit 4 L (nad4L) in *A. mecnuffi*, which was also revealed to be an error solved after manual curation. Because no reference mitogenomes from the same genus were available for comparison, the accuracy of the SR assemblies was evaluated indirectly. Their complete circularisation, negligible intraspecific length variation and successful annotation of all genes usually present in tardigrade mitogenomes supported their interpretation as the most accurate assemblies. The five mitogenomes are available on GenBank under accession numbers PZ148665-9.

3.3 | Palindromes Quantification in Reads

Palindrome quantification was performed by aligning the long reads used for the LR assembly (i.e., only reads matching the mitochondrial proteins database) against their respective mitogenome obtained through SR assembly. Reads with multiple matches in different orientations to their respective genomes were considered as palindromic. The majority of the reads had a blastn match to their mitogenome (Table 3), and about half of the reads selected for the assembly by blastx against a mitochondrial database, which matched the final mitogenome assemblies (48.39%–56.00%; Table 3), had only one match. Among the reads with at least two matches to the mitogenome, more than two-thirds (68.41%–74.51%; Table 3) were identified as palindromic. Examples of the structure of those palindromic reads are presented in Figure 2.

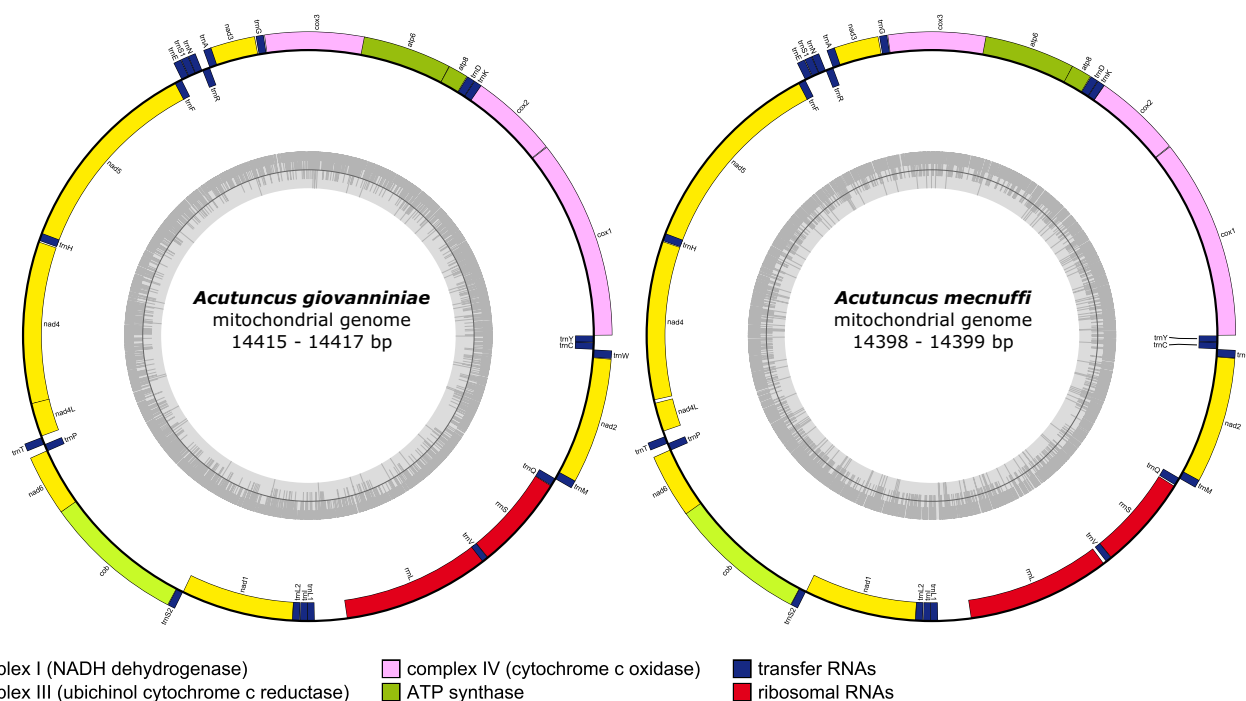


FIGURE 1 | Graphical representation of mitochondrial genomes of *A. giovanninae* and *A. mecnuffi*.

TABLE 3 | Results of palindrome quantification in reads through blastn against the corresponding assembled mitogenome.

Individual	# of reads	# of reads matched to mitogenome	# of reads with one match to the mitogenome	# of palindromic reads	% reads with one match to the mitogenome	% of palindromic reads (over all matched reads)	% of palindromic reads (over reads with at least two matches to the mitogenome)
Acu_gio_RIT094	30,888	27,103	13,114	10,423	48.39	38.46	74.51
Acu_gio_RPL027	535,322	522,167	264,184	178,060	50.59	34.1	69.02
Acu_gio_RSE010	1,840,681	1,810,510	859,350	653,968	47.46	36.12	68.75
Acu_mes_RSE086	125,643	109,818	61,499	33,365	56.00	30.38	69.05
Acu_mes_RSE087	32,919	22,398	14,028	5726	62.63	25.56	68.41

4 | Discussion

The introduction of WGA has been a key technological innovation, enabling the genomic characterisation of microscopic taxa for which the limiting factor for a long time was low DNA content. We documented and evaluated how the use of MDA, which, due to its high reaction speed and amplification gain, introduces palindromic artefacts into mitochondrial genomes, affects mitogenome assembly after long-read sequencing and how this issue can be mitigated. We demonstrated that, while long-read assemblies based on MDA frequently fail in the presence of pervasive palindromes, simple in silico fragmentation of long reads into short, high-quality fragments enhances robust and reproducible mitogenome reconstruction.

The application of MDA to single microscopic creatures has facilitated their genome and mitogenome sequencing, but

the formation of palindromic reads still represents an issue. In our study system, roughly one-third of all mitochondrial reads were palindromic (Table 3), that is, matches to the mitochondrial genomes in different directions were present. This proportion of palindromic reads is comparable with other similar studies (Lu et al. 2023), but the complex structure of palindromic repeats found in this study (Figure 2) was not reported before. Palindromic read generation can potentially be explained by the high reaction gain of the REPLI-g Advanced Single Cell kit used for this study (which can yield up to 35 µg of the amplified product from a single cell). It could be hypothesised that the high reaction gains and reaction speed (2h) could increase the probability of multiple primers annealing with the template in the same region, with consequent strand displacement and palindrome formation. To support this association between reaction speed and palindrome formation, two previous studies (Camarda et al. 2025; Vecchi

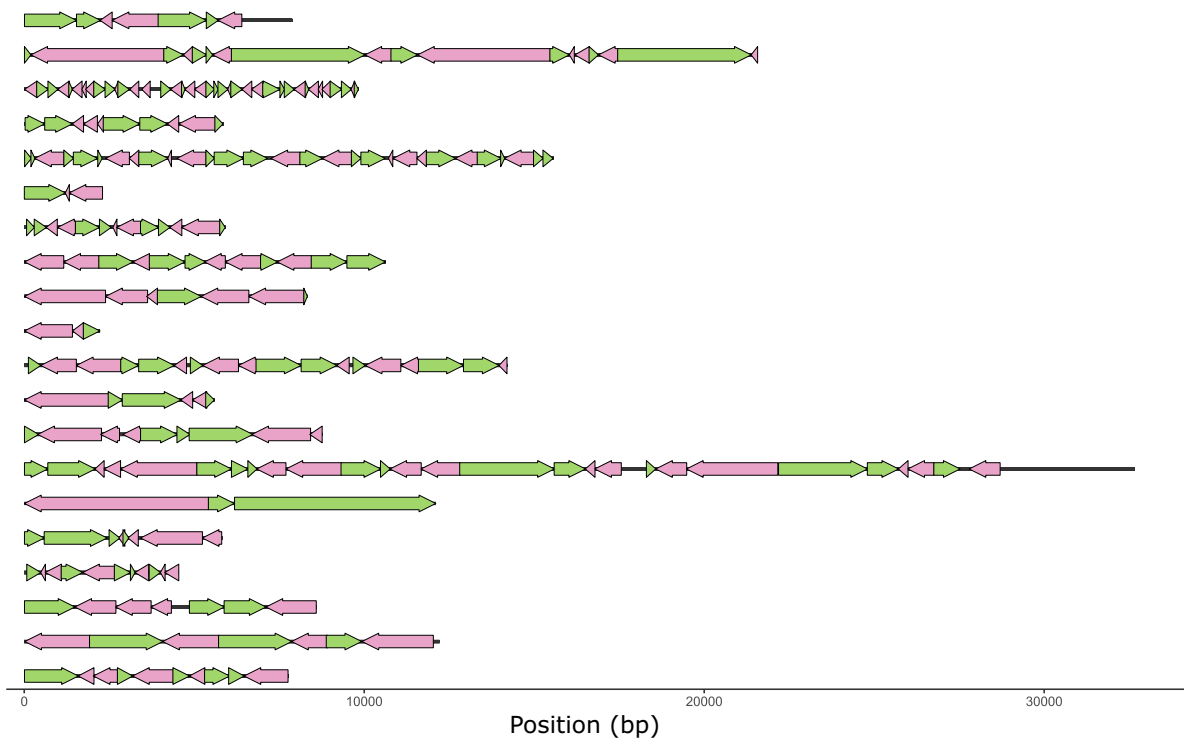


FIGURE 2 | Example of palindromic reads (random sample). Arrows indicate blastn matches to reference mitogenomes according to match orientation. Colour highlights alternative matching directions.

and Stec 2025) did not find any assembly issues on tardigrade mitogenomes while using an almost identical protocol to this study. The main difference was the use of a REPLI-g Mini Kit (Qiagen), which has lower yield and a much longer reaction time (10–16 h vs. 2 h in the REPLI-g Advanced Single Cell kit). Similar amplification-associated palindrome formation has been documented in long-read datasets derived from single nematodes (Lee et al. 2023) and other microscopic organisms (Agyabeng-Dadzie et al. 2025), but it has never been precisely quantified. Our findings therefore provide empirical evidence that palindromic read formation is not a marginal artefact but can be a dominant feature of high-gain and high-speed MDA in mitogenomes.

4.1 | Long-Read Assembly in the Presence of Pervasive Palindromes

Assembly attempts directly on reads with a high proportion of palindromes have proven mostly unsuccessful (Table 3). Long-read assemblers such as Flye rely on repeat graph construction and assume that read structure reflects contiguous genomic segments (Lin et al. 2016). Extensive palindromic inversions violate this assumption, artificially introducing inverted repeats and structural complexity that can fragment assembly graphs or promote misassembly. Bioinformatic tools to correct MDA-induced palindromes in long reads have already been devised, and after identifying palindromic reads, they either discard them (Agyabeng-Dadzie et al. 2025) or split them at detected inverted alignments (Lee et al. 2023; Lu et al. 2023; Warris et al. 2018). While the first approach can result in a non-negligible data loss when palindromic reads are present at high abundance, the second approach mostly works when palindromic repeats are long (above a user-specified

threshold) and only one or two inversion events are present in each palindromic read. In our data, most of the palindromic reads contain complex multi-segment structures (Figure 2), suggesting repeated strand-displacement and re-priming events and effectively preventing mitogenome assembly. Moreover, the two mentioned correction approaches rely primarily on aligning the reads to their reverse complement and flagging them as palindromic if matches above a certain threshold are found. However, these methods are computationally intensive as they require an additional mapping or alignment step.

4.2 | Read Fragmentation as a Robust Solution to Mitogenome Assembly

While palindrome detection and correction have been shown to be effective for nuclear genomes, a different strategy proved more suitable for mitochondrial genomes. In silico fragmentation of long reads into 150 bp high-quality fragments, followed by short-read assembly with NOVOPlasty, consistently yielded circular mitogenomes of the expected size with negligible and biologically plausible length variation across individuals. Importantly, coverage values in the short-read assemblies were several orders of magnitude higher than those obtained with long-read assemblies. While in our case, due to the high sequencing depth, coverage was not an issue, in some cases, when economical constraints limit the number of obtainable reads, the higher efficiency of the proposed methods provides clear advantages. The fragmentation eliminates the structural artefacts by breaking the long palindromic sequences into segments that represent the true mitochondrial sequences, but without the problematic and erroneous structural information. Only a small number of short fragments will retain the junctions between inverted repeats;

the successful assemblies show that those still problematic reads are not numerous enough to cause issues for assemblers. While long-range spatial information is lost, it is inconsequential when dealing with small circular genomes (~14.5 kb) that can easily be assembled from short overlapping fragments. As mitochondrial genomes are circular, small and in many cases lack extensive repeats, they are especially amenable to this strategy. For nuclear genome assembly, however, fragmentation would sacrifice essential long-range information and would not constitute a viable solution. Overall, the proposed read-fragmentation approach provides a practical and robust alternative to previously existing tools for palindrome removal (Lee et al. 2023; Lu et al. 2023; Warris et al. 2018) for the assembly of small genomes that naturally lack extensive repeated sequences, where there is less need for the long-range spatial information provided by long reads and in which MDA has produced complex palindromic structures.

4.3 | Applications, Limitations and Future Directions

From the results presented here, it would seem advantageous to sequence MDA amplified libraries with short-read technologies instead of long-reads (like Nanopore), as long reads would need to be fragmented anyway before mitogenome assembly. However, Nanopore sequencing platforms offer practical advantages over commercial short-read platforms: devices are relatively inexpensive and scalable, allowing sequencing to be performed in small laboratories or even field environments, and sequencing runs can be monitored in real time to stop once sufficient coverage has been obtained (Menegon et al. 2017; Parker et al. 2017). Information provided by long reads could still be used for the assembly of nuclear genomes, for which already available palindrome-correction tools have been shown to be effective (Lee et al. 2023). Even when short-read assembly strategies are used, generating data through Nanopore sequencing remains a powerful and versatile approach.

An important limitation of this approach lies in the inherently high indel error rate of Nanopore sequencing compared to Illumina (Chen et al. 2025). In our approach, the fragmented Nanopore reads are selected for quality $Q > 30$; however, the error profile could still be different than true Illumina reads, in particular regarding indels. This could lead to errors in the mitogenome assembly by software designed specifically for Illumina sequences, and therefore, assembled mitogenomes must be carefully inspected, and particular care should be given to check for the presence of complete ORFs, which would be easily disrupted by indels introduced in the assembly process.

While the current study used tardigrades as a model to develop the proposed mitogenome assembly pipeline, the approach is extremely flexible regarding taxonomy: any dataset derived from long-read sequencing of MDA amplified genomes has the potential to result in the formation of palindromic reads, which could then benefit from the approach proposed here when assembling the mitogenome. The main purpose of read pre-filtering based on a mitochondrial protein database in this study was to reduce the number of reads to process; for sequencing runs with lower outputs, this step may not be needed at all. For sequencing runs with lower output, this step may not be needed. If pre-filtering is required, the taxonomic limitation imposed by a reference

database can be reduced by using a custom database generated for the organism or group of interest.

Also, the precise detection and quantification of palindromic reads require mapping them to a reference genome (Lu et al. 2023), which in our case was easy to obtain, but which is not always the case with the bigger and more complex nuclear genomes. Published research shows that in nematodes the presence of palindromes in nuclear reads can be easily solved with the palindromic reads' detection and correction bioinformatic tools mentioned above (Lee et al. 2023). However, in this study, the authors used an MDA kit with lower gain and reaction speed; thus, the outcome cannot be easily compared with our results. Future research should try to quantify the influence of diverse MDA kits and genomic DNA input characteristics on the production of palindromic sequences.

To summarise, this study showed that high reaction-gain and high-speed MDA applied to low-biomass specimens can generate pervasive palindromic reads that severely impair long-read mitogenome assembly. However, simple in silico fragmentation of long reads into artificial short reads enables consistent and accurate reconstruction of circular mitochondrial genomes. This strategy provides a practical and scalable solution for mitogenomics of microscopic animals and supports broader efforts to reduce taxonomic bias in genomic resources.

Author Contributions

Matteo Vecchi: conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualisation, writing – original draft, writing – review and editing. **Bartłomiej Surmacz:** validation, writing – review and editing. **Ingemar K. Jönsson:** resources, writing – review and editing. **Daniel Stec:** conceptualisation, project administration, supervision, writing – review and editing.

Acknowledgements

We wish to thank Valentina Peona (Swedish Museum of Natural History, Stockholm, Sweden) and Sara Brandoli (Ca' Foscari University of Venice, Italy) for their critical reading of the manuscript. This study was funded by the Polonez Bis grant No. 2022/45/P/NZ8/01512 to M.V. co-funded by the National Science Centre and the European Union framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie grant agreement [No. 945339]. Samples were collected under sampling permits: N.0003526/2021 of 30/11/2021 (Parco dell'Appennino Tosco-Emiliano, Italy), N.798 of 08/2023 (Karkonoski Park Narodowy, Poland) and N. 521-2330-2023 of 23/05/2023 (Länsstyrelsen Blekinge, Sweden).

Funding

This study was funded by the Polonez Bis grant No. 2022/45/P/NZ8/01512 to M.V. co-funded by the National Science Centre and the European Union framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie grant agreement [No. 945339].

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All the data and code produced in this study are available as [Supporting Information](#) of this article (Code used for the analyses: SM.01) and/or deposited in open-access databases (Raw reads: NCBI SRA BioProject PRJNA1287536; Assembled mitogenomes: GenBank PZ148665-9; Code used for the analyses: <https://doi.org/10.5281/zenodo.19131525> and SM.01).

References

- Agyabeng-Dadzie, F., M. S. Beaudry, A. Deyanov, et al. 2025. "Evaluating the Benefits and Limits of Multiple Displacement Amplification With Whole-Genome Oxford Nanopore Sequencing." *Molecular Ecology Resources* 25, no. 6: e14094. <https://doi.org/10.1111/1755-0998.14094>.
- Bemm, F., C. L. Weiß, J. Schultz, and F. Förster. 2016. "Genome of a Tardigrade: Horizontal Gene Transfer or Bacterial Contamination?" *Proceedings of the National Academy of Sciences* 113, no. 22: E3054–E3056. <https://doi.org/10.1073/pnas.1525116113>.
- Benita, Y. 2003. "Regionalized GC Content of Template DNA as a Predictor of PCR Success." *Nucleic Acids Research* 31, no. 16: 99e–999e. <https://doi.org/10.1093/nar/gng101>.
- Bernt, M., A. Donath, F. Jühling, et al. 2013. "MITOS: Improved de novo metazoan mitochondrial genome annotation." *Molecular Phylogenetics and Evolution* 69, no. 2: 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>.
- Bonenfant, Q., L. Noé, and H. Touzet. 2023. "Porechop_ABI: Discovering Unknown Adapters in Oxford Nanopore Technology Sequencing Reads for Downstream Trimming." *Bioinformatics Advances* 3, no. 1: vbac085. <https://doi.org/10.1093/bioadv/vbac085>.
- Borgström, E., M. Paterlini, J. E. Mold, J. Frisen, and J. Lundeberg. 2017. "Comparison of Whole Genome Amplification Techniques for Human Single Cell Exome Sequencing." *PLoS One* 12, no. 2: e0171566. <https://doi.org/10.1371/journal.pone.0171566>.
- Buchfink, B., K. Reuter, and H. G. Drost. 2021. "Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND." *Nature Methods* 18, no. 4: 366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- Camacho, C., G. Coulouris, V. Avagyan, et al. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10, no. 1: 421.
- Camarda, D., O. Lisi, D. Stec, and M. Vecchi. 2025. "Description of a New Genus and Species of Isohypsibioidea (Tardigrada), Together With Its Mitochondrial Genome Sequence." *Arthropod Systematics & Phylogeny* 83: 427–445. <https://doi.org/10.3897/asp.83.e150460>.
- Chen, S., F. Nie, and J. Wang. 2025. "NanoFilter: Enhancing Phasing Performance by Utilizing Highly Consistent INDELs and SNVs in Nanopore Sequencing." *Bioinformatics* 41, no. 9: btaf453. <https://doi.org/10.1093/bioinformatics/btaf453>.
- Dean, F. B., S. Hosono, L. Fang, et al. 2002. "Comprehensive Human Genome Amplification Using Multiple Displacement Amplification." *Proceedings of the National Academy of Sciences* 99, no. 8: 5261–5266. <https://doi.org/10.1073/pnas.0820894999>.
- Dierckxsens, N., P. Mardulyn, and G. Smits. 2016. "NOVOPlasty: de novo Assembly of Organelle Genomes From Whole Genome Data." *Nucleic Acids Research* 28, no. 45: e18. <https://doi.org/10.1093/nar/gkw955>.
- Donath, A., F. Jühling, M. Al-Arab, et al. 2019. "Improved Annotation of Protein-Coding Genes Boundaries in Metazoan Mitochondrial Genomes." *Nucleic Acids Research* 47, no. 20: 10543–10552. <https://doi.org/10.1093/nar/gkz833>.
- Flot, J.-F., B. Hespeels, X. Li, et al. 2013. "Genomic Evidence for Ameiotic Evolution in the Bdelloid Rotifer *Adineta vaga*." *Nature* 500, no. 7463: 453–457. <https://doi.org/10.1038/nature12326>.
- Fontaneto, D., J.-F. Flot, and C. Q. Tang. 2015. "Guidelines for DNA Taxonomy, With a Focus on the Meiofauna." *Marine Biodiversity* 45, no. 3: 433–451. <https://doi.org/10.1007/s12526-015-0319-7>.
- Forni, G., F. Plazzi, A. Cussigh, et al. 2021. "Phylomitogenomics Provides New Perspectives on the Euphasmatodea Radiation (Insecta: Phasmatodea)." *Molecular Phylogenetics and Evolution* 155: 106983. <https://doi.org/10.1016/j.ympev.2020.106983>.
- Gammuto, L., V. Serra, G. Petroni, and M. A. Todaro. 2024. "Molecular Phylogenetic Position and Description of a New Genus and Species of Freshwater Chaetonotidae (Gastrotricha: Chaetonotida: Paucitubulatina), and the Annotation of Its Mitochondrial Genome." *Invertebrate Systematics* 38, no. 7: IS23059. <https://doi.org/10.1071/IS23059>.
- García-Souto, D., S. Zumalave, J. M. Martínez-Romero, N. Marí-Mena, A. Vizcaino, and M. Vila. 2025. "Phylomitogenomics Supports *Actias Isabellae* (Graells, 1849) as the Definitive Scientific Name of the Spanish Moon Moth (Lepidoptera, Saturniidae)." *Genetica* 153, no. 1: 15. <https://doi.org/10.1007/s10709-025-00231-w>.
- Greiner, S., P. Lehwark, and R. Bock. 2019. "OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes." *Nucleic Acids Research* 47, no. W1: W59–W64. <https://doi.org/10.1093/nar/gkz238>.
- Huang, S., M. Kang, and A. Xu. 2017. "HaploMerger2: Rebuilding Both Haploid Sub-Assemblies From High-Heterozygosity Diploid Genome Assembly." *Bioinformatics* 33, no. 16: 2577–2579. <https://doi.org/10.1093/bioinformatics/btx220>.
- Kim, B. Y., H. R. Gellert, S. H. Church, et al. 2024. "Single-Fly Genome Assemblies Fill Major Phylogenomic Gaps Across the Drosophilidae Tree of Life." *PLoS Biology* 22, no. 7: e3002697. <https://doi.org/10.1371/journal.pbio.3002697>.
- Kocot, K. M., T. H. Struck, J. Merkel, et al. 2017. "Phylogenomics of Lophotrochozoa With Consideration of Systematic Error." *Systematic Biology* 66, no. 2: 256–282. <https://doi.org/10.1093/sysbio/syw079>.
- Kolmogorov, M., D. M. Bickhart, B. Behsaz, et al. 2020. "metaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs." *Nature Methods* 17, no. 11: 1103–1110. <https://doi.org/10.1038/s41592-020-00971-x>.
- Kosakyan, A., L. Gammuto, A. Cesaretti, et al. 2026. "From Rigid Order to Radical Variation: Mitogenome Evolution in the Main Lineages of a Lesser-Known Animal Phylum (Gastrotricha)." *Genome Biology and Evolution* 18, no. 2: evag001. <https://doi.org/10.1093/gbe/evag001>.
- Lasken, R. S., and T. B. Stockwell. 2007. "Mechanism of Chimera Formation During the Multiple Displacement Amplification Reaction." *BMC Biotechnology* 7, no. 1: 19. <https://doi.org/10.1186/1472-6750-7-19>.
- Laumer, C. E., R. Fernández, S. Lemer, et al. 2019. "Revisiting Metazoan Phylogeny With Genomic Sampling of All Phyla." *Proceedings of the Royal Society B: Biological Sciences* 286, no. 1906: 20190831. <https://doi.org/10.1098/rspb.2019.0831>.
- Lavrov, D. V., M. C. Diaz, M. Maldonado, et al. 2023. "Phylomitogenomics Bolsters the High-Level Classification of Demospongiae (Phylum Porifera)." *PLoS One* 18, no. 12: e0287281. <https://doi.org/10.1371/journal.pone.0287281>.
- Lee, Y.-C., H.-M. Ke, Y.-C. Liu, et al. 2023. "Single-Worm Long-Read Sequencing Reveals Genome Diversity in Free-Living Nematodes." *Nucleic Acids Research* 51, no. 15: 8035–8047. <https://doi.org/10.1093/nar/gkad647>.
- Li, L., Z. Ge, S. Liu, et al. 2024. "Multi-Omics Landscape and Molecular Basis of Radiation Tolerance in a Tardigrade." *Science* 386, no. 6720: ead10799. <https://doi.org/10.1126/science.adl0799>.
- Lin, Y., J. Yuan, M. Kolmogorov, M. W. Shen, M. Chaisson, and P. A. Pevzner. 2016. "Assembly of Long Error-Prone Reads Using de Bruijn Graphs." *Proceedings of the National Academy of Sciences* 113, no. 52: E8396–E8405. <https://doi.org/10.1073/pnas.1604560113>.

- Lu, N., Y. Qiao, P. An, et al. 2023. "Exploration of Whole Genome Amplification Generated Chimeric Sequences in Long-Read Sequencing Data." *Briefings in Bioinformatics* 24, no. 5: bbad275. <https://doi.org/10.1093/bib/bbad275>.
- Mammola, S., M. Adamo, D. Antić, et al. 2023. "Drivers of Species Knowledge Across the Tree of Life." *eLife* 12: RP88251. <https://doi.org/10.7554/eLife.88251.3>.
- Martínez, A., S. Bonaglia, M. Di Domenico, et al. 2025. "Fundamental Questions in Meiofauna Research Highlight How Small but Ubiquitous Animals Can Improve Our Understanding of Nature." *Communications Biology* 8, no. 1: 449. <https://doi.org/10.1038/s42003-025-07888-1>.
- Menegon, M., C. Cantaloni, A. Rodriguez-Prieto, et al. 2017. "On Site DNA Barcoding by Nanopore Sequencing." *PLoS One* 12, no. 10: e0184741. <https://doi.org/10.1371/journal.pone.0184741>.
- Parker, J., A. J. Helmstetter, D. Devey, T. Wilkinson, and A. S. T. Papadopoulos. 2017. "Field-Based Species Identification of Closely-Related Plants Using Real-Time Nanopore Sequencing." *Scientific Reports* 7, no. 1: 8345. <https://doi.org/10.1038/s41598-017-08461-5>.
- Pollard, M. O., D. Gurdasani, A. J. Mentzer, T. Porter, and M. S. Sandhu. 2018. "Long Reads: Their Purpose and Place." *Human Molecular Genetics* 27, no. R2: R234–R241. <https://doi.org/10.1093/hmg/ddy177>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Roberts, N. G., M. J. Gilmore, T. H. Struck, and K. M. Kocot. 2024. "Multiple Displacement Amplification Facilitates SMRT Sequencing of Microscopic Animals and the Genome of the Gastrotrich *Lepidodermella squamata* (Dujardin 1841)." *Genome Biology and Evolution* 16, no. 12: evae254. <https://doi.org/10.1093/gbe/evae254>.
- Sabina, J., and J. H. Leamon. 2015. "Bias in Whole Genome Amplification: Causes and Considerations." In *Whole Genome Amplification*, edited by T. Kroneis, vol. 1347, 15–41. Springer New York. https://doi.org/10.1007/978-1-4939-2990-0_2.
- Sayers, E. W., J. Beck, E. E. Bolton, et al. 2025. "Database Resources of the National Center for Biotechnology Information in 2025." *Nucleic Acids Research* 53, no. D1: D20–D29. <https://doi.org/10.1093/nar/gkae979>.
- Shen, W., S. Le, Y. Li, and F. Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation." *PLoS One* 11, no. 10: e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
- Simion, P., K. Belkhir, C. François, et al. 2018. "A Software Tool 'CroCo' Detects Pervasive Cross-Species Contamination in Next Generation Sequencing Data." *BMC Biology* 16, no. 1: 28. <https://doi.org/10.1186/s12915-018-0486-7>.
- Sun, X., D. Yu, Z. Xie, et al. 2020. "Phylomitogenomic Analyses on Collembolan Higher Taxa With Enhanced Taxon Sampling and Discussion on Method Selection." *PLoS One* 15, no. 4: e0230827. <https://doi.org/10.1371/journal.pone.0230827>.
- Vecchi, M., C. Ferrari, D. Stec, and S. Calhim. 2022. "Desiccation Risk Favours Prevalence and Diversity of Tardigrade Communities and Influences Their Trophic Structure in Alpine Ephemeral Rock Pools." *Hydrobiologia* 849, no. 9: 1995–2007. <https://doi.org/10.1007/s10750-022-04820-0>.
- Vecchi, M., C. Ferrari, D. Fontaneto, J. Godziek, I. K. Jönsson, and D. Stec. 2026. "DNA Metabarcoding Reveals Tardigrades' Diversity and Dispersal Patterns in European Freshwater Rock Pools." *Journal of Biogeography* 53, no. 1: e70128. <https://doi.org/10.1111/jbi.70128>.
- Vecchi, M., and D. Stec. 2025. "Mitogenome of a New Ramazzottius Species (Tardigrada: Eutardigrada: Ramazzottiidae) Discovered in Rock Pools Along With Its Temperature and Desiccation-Related Proteins Repertoire." *Organisms Diversity & Evolution* 25: 119–135. <https://doi.org/10.1007/s13127-024-00662-x>.
- Vecchi, M., A. Tsvetkova, D. Stec, C. Ferrari, S. Calhim, and D. Tumanov. 2023. "Expanding Acutuncus: Phylogenetics and Morphological Analyses Reveal a Considerably Wider Distribution for This Tardigrade Genus." *Molecular Phylogenetics and Evolution* 180: 107707. <https://doi.org/10.1016/j.ympev.2023.107707>.
- Wang, X., Y. Liu, H. Liu, et al. 2022. "Recent Advances and Application of Whole Genome Amplification in Molecular Diagnosis and Medicine." *MedComm* 3, no. 1: e116. <https://doi.org/10.1002/mco2.116>.
- Warris, S., E. Schijlen, H. Van De Geest, et al. 2018. "Correcting Palindromes in Long Reads After Whole-Genome Amplification." *BMC Genomics* 19, no. 1: 798. <https://doi.org/10.1186/s12864-018-5164-1>.
- Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, H., M. Averick, J. Bryan, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4, no. 43: 1686. <https://doi.org/10.21105/joss.01686>.
- Zardoya, R., and M. Suárez. 2008. "Sequencing and Phylogenomic Analysis of Whole Mitochondrial Genomes of Animals." In *Phylogenomics*, edited by W. J. Murphy, vol. 422, 185–200. Humana Press. https://doi.org/10.1007/978-1-59745-581-7_12.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supporting Information: S1.** Code used in the manuscript analyses.