

[CC BY 4.0 DEED Attribution 4.0 International](#)

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Draft genome from facultatively parthenogenetic Opiliones indicates frequent mitonuclear sequence transfer and novel full-length insertions

Sarah Stellwagen

University of North Carolina at Charlotte

Mercedes Burns (✉ burnsm@umbc.edu)

University of Maryland, Baltimore County

Research Article

Keywords: genome assembly, Opiliones, polyploidy, long-read sequencing, facultative parthenogenesis

Posted Date: January 11th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-3846124/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Draft genome from facultatively parthenogenetic Opiliones indicates frequent mitonuclear sequence transfer and novel full-length insertions

Sarah Stellwagen^{1,2,*} and Mercedes Burns^{2,†}

¹Department of Biological Sciences, University of North Carolina at Charlotte and ²Department of Biological Sciences, University of Maryland, Baltimore County

January 9, 2024

Abstract

Background: Facultative parthenogenesis and intra-population mixed ploidy are rare in animals. These unique characteristics allow opportunities to investigate the relationship between sexual modality and ploidy. We have completed a draft genome of the Japanese harvester ("daddy-longlegs") *Leiobunum manubriatum*, a species which reproduces sexually and asexually, and with mixed diploid and tetraploid populations in some areas. **Results:** We combined Oxford Nanopore's MinION long-read sequencing platform with Dovetail Hi-C scaffolding to assemble the haploid genome for the diploid race, which is approximately 336 MBp after collapsing heterozygous sequence. The assembly's completeness was measured using BUSCOs from Eukaryota (complete: 92.6%), Arthropoda (complete: 96.9%), and Arachnida (complete: 95.3%). We searched raw sequence reads and the draft genome for nuclear mitochondrial DNA (numt) sequences. While only one complete mitochondrial genomic transfer was found in the draft genome, there are at least 12 complete numts across 9 reads within the raw sequencing data that were lost during the assembly process. **Conclusions:** The genome of the *L. manubriatum* diploid race is an invaluable resource not only for opiloid research, but also for facilitating studies investigating the evolution of their unique reproductive mode and mixed ploidy. To our knowledge, this is the first published genome of a wild-derived facultative parthenogen. Future work will leverage this resource in comparative genomics and transcriptomics of *L. manubriatum* to understand the connection between ploidy and sexual strategy.

Key words: genome assembly; Opiliones; polyploidy; long-read sequencing; facultative parthenogenesis

Introduction

Sexual reproduction is the dominant mode of reproduction, however many animals and plants exhibit parthenogenesis, or the production of viable offspring without fertilization of an egg. Particularly in plants and arthropods, approximately one out of every 1000 multicellular eukaryotic taxa exhibit parthenogenesis [1]. Polyploidy, the condition where an organism has more than two sets of chromosomes, is often associated with asexuality [2, 3]. The *Daphnia pulex* complex of water fleas has both obligate asexual and cyclically sexual/asexual populations, in addition to polyploidization of some lineages [4]. Populations of New Zealand freshwater snail *Potamopyrgus antipodarum*, which include sexual and asexual lineages, have variations in ploidy level and genome size [5]. The spiny leaf insect *Extatosoma tiaratum* exhibits facultative parthenogenesis, where females can produce offspring asexually, or, if mated, sexually [6].

Two species of Japanese opiloidids (also known as "harvesters" or "daddy-longlegs") belonging to the *Leiobunum curvipalpe* group exhibit facultative parthenogenesis: *L. manubriatum* (Figure 1) and *L. globosum*. These species are endemic to northern Japan, with *L. manubriatum*'s range extending through the Japanese Alps and overlapping on the range of *L. globosum* in Aomori, Akita, and Hokkaido Prefectures. While *L. globosum* maintains tetraploidy in all individuals [7], those of *L. manubriatum* have intra-population diploidy and tetraploidy, or a single cytotype in some populations [8]. Both ploidies can reproduce sexually or asexually, but it is not currently known whether mated females can produce unfertilized eggs even after mating. The number of chromosomes for the mixed-sex diploid race of *L. manubriatum* is reported to be $2n=24$, with the all-female tetraploid race varying between $2n=4x=ca. 48$ (46–49) [7].

Nuclear mitochondrial sequence (numt) refers to areas of the nuclear genome that contains sequence which originated in the mitochondrial genome [9]. Numts tend to be relatively fragmented, and to our knowledge, there is only one reported full-length mitogenome insertion to the nuclear genome, discovered in tarsiers [10]. This insertion covered the complete 17,004 bp



Figure 1. Adult *Leiobunum manubriatum* male (left) and female from Shomyo Falls, Japan, copulating. Photo credit: Sarah Stellwagen

mitogenome with an additional 862 bp overlap of the D-loop region and partial 16S rRNA [11, 10]. The largest human numt covers approximately 90% of the mitochondrial genome [12]. In arthropods, honeybees have the highest percentage of reported numts, though their numts are relatively short, the longest being 3,335 bp and spread across a 25 kb area of the nuclear genome [13].

Mitochondrial genomes average around 17 kb [14], and duplications can confuse assembly efforts when aligning sequences obtained through short read technologies [15, 16]. Furthermore, nuclear insertions of mitochondrial sequence can lead to problems with phylogenetic reconstruction if numts are inadvertently amplified [17] or present in only a subset of taxa. Long-read sequencing, a necessary tool for overcoming the challenges of repetitive genomics, has exploded on the scene, however there are only a few studies that have used long-read sequencing to study mitochondrial biology. These studies have used this new technology to gain insight into heteroplasmy and mitogenome rearrangement [18], mitochondrial DNA variant analysis [19], and new computational pipelines to assemble mitochondrial genomes [20].

Here, we describe the *de novo* sequencing and assembly of the genome of the diploid race of harvester species *L. manubriatum* using the long-read sequencing platform from Oxford Nanopore Technologies combined with Dovetail scaffolding. This draft genome is the first of a facultatively parthenogenetic harvester species, and the second of Opiliones following the genome of the widely distributed, human-associated species *Phalangium opilio*, [21]. The *L. manubriatum* nuclear genome has unusually large numt insertions and documents the first incidence of multiple full-length transfers in animals. Facultative parthenogenesis is rare in organisms, but provides an interesting case study on the benefits and consequences for the evolution of sex. The genome assembled is moreover singular in that it is, to our knowledge, the first published genome of a wild-derived facultative parthenogen (but see [22] and [23] for captive-bred examples). *L. manubriatum* presents a unique combination of polyploidy and facultative asexuality, and understanding the complexity of their genomic make-up will allow a deeper insight into the maintenance of sex.

Methods

Sample Collection and Extraction

Adult female *L. manubriatum* specimens were collected from forest around Hirayu Campground (Figure 2) on July 11, 2014 and August 3, 2019 and Shōmyo Falls (Figure 2) visiting area on July 11, 2014 and August 3–4, 2019. Specimens collected in 2014 were stored in 100% ethanol. Specimens collected in 2019 were immediately transported live to Tottori University, Tottori, Japan for DNA extraction. In order to reduce the amount of contaminating DNA from gut flora and parasites (e.g. gregarines), the gut of each specimen was dissected and removed. High molecular weight DNA was then extracted from the remaining tissue of each specimen using the MasterPure Complete DNA and RNA Purification Kit following the DNA Purification section protocol (cat.no. MC89010). DNA samples were then transported to the University of Maryland, Baltimore County, Maryland, USA for further processing and sequencing.

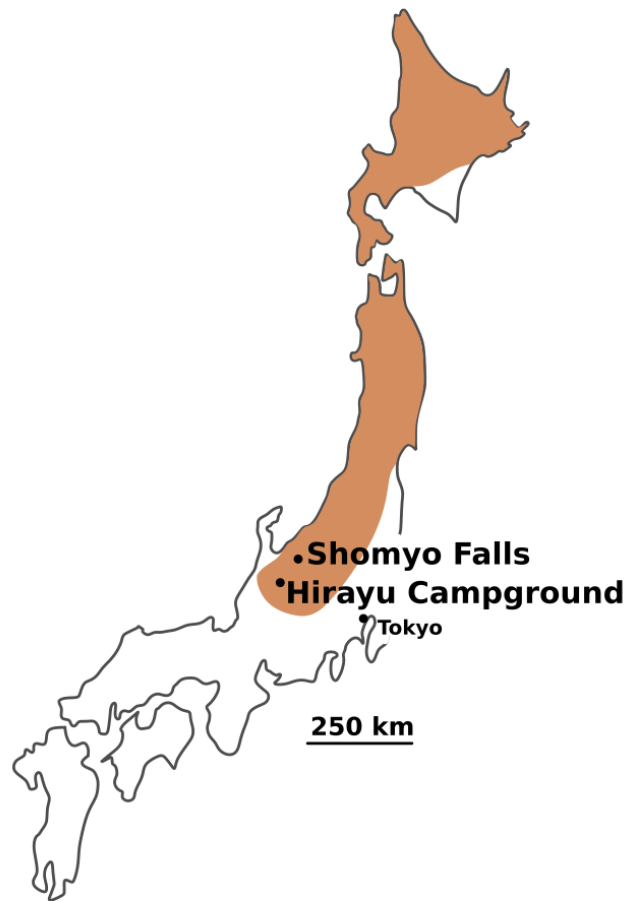


Figure 2. Map of Japan showing *L. manubriatum* distribution (tan shading) and collecting sites (Shōmyo Falls and Hirayu Campground).

Table 1. *Leiobunum manubriatum* nuclear genome assembly statistics. BUSCO scores are complete (single plus duplicated).

| Genome Assembly | Value |
|-------------------------------------|----------------|
| Nanopore Sequencing Statistics | |
| Number of Reads (Q10) | 7,958,356 |
| Number of Bases (bp) | 46,760,569,792 |
| Assembly Statistics | |
| Assembly Size (bp) | 336,872,803 |
| %CG | 37.48 |
| Number of Contigs | 3,399 |
| Longest Contig (bp) | 71,836,609 |
| N50(bp) | 27,489,741 |
| L50(bp) | 4 |
| Protein-coding genes | 24,032 |
| BUSCO Scores | |
| Assembly, Arthropoda | 96.9% |
| Annotation, Arthropoda, transcripts | 93.6% |
| Annotation, Arthropoda, proteins | 93.6% |

Nanopore Sequencing

The DNA from 29 specimens was used for sequencing. Extracted DNA was combined to reach 10 ug pooled samples and loaded onto a Sage Science BluePippin cassette (cat.no. BLF7150 or BPLUS10) and run with a 10, 15, or 20 kbp high pass threshold overnight, or prepped without size selection. The resultant samples were then cleaned using Agencourt AMPure XP beads (cat.no. A63881) and eluted overnight to several days in water. Clean DNA was then used in Oxford Nanopore's 1D Genomic DNA by Ligation protocol (SQK-LSK109). A total of 11 runs were completed using SpotON Flow Cells (R9.4; cat.no. FLO-MIN106) and the resultant fast5 files were basecalled using Oxford Nanopore's program Guppy 3.4.4+a296acb, and filtered to include only those with a Q-score of 10 or higher. Adapter sequences were then trimmed using Porechop v0.2.4 (Porechop, RRID:SCR_016967).

De novo Nuclear Genome Assembly

Trimmed reads were assembled using Canu v1.9 (Canu, RRID:SCR_015880) [24] with default parameters. The raw draft assembly was then further scaffolded by Dovetail HiRise. The resulting draft assembly was then polished using Nanopore's Medaka v1.0.1 program. Purge Haplotigs v1.1.1 [25] was then used on the polished assembly to remove heterozygous haplotype contigs that were assembled separately with $a=50$.

The final assembly is 336 Mbp from 3,399 contigs, with an N50 of 27,489,741 bp (Table 1; NCBI Project: PRJNA814647). Half of the genome is represented by 4 contigs (L50). The genome recovered 96.9% of the 1,066 BUSCO [26] arthropod genes (Table 1; Single: 94.7.8%, Duplicated: 2.2%, Fragmented: 0.8%, Missing: 2.3%). These BUSCO scores are excellent compared to recent spider assemblies, for example the chromosome-scale *Argiope bruennichi* genome, which used Illumina, PacBio, and Hi-C sequencing, recovered 91.1% complete arthropod BUSCOs [27]. The *Dysdera silvatica* genome, which used Illumina paired end and mate pair sequencing in addition to both PacBio and Oxford Nanopore sequencing, recovered 69.1% complete BUSCOs [28]. While chromosome scale genome organization is not currently feasible with Oxford Nanopore alone, this sequencing strategy can outperform completeness estimates compared to mixing various technologies that achieve chromosome scale resolution.

The polishing program Medaka (Oxford Nanopore) combined with Purge Haplotigs [25] to reduce heterozygous haplotype contigs greatly improved BUSCO completeness metrics, while reducing duplications (Figure 1). The first round of Medaka polishing increased complete BUSCOs by nearly 10%, however duplications also increased by 4%. Fragmented and missing BUSCOs were also greatly reduced, and only slightly increased (<1%) after purging haplotigs. Purge Haplotigs greatly duplicated BUSCOs by over 10% without severely affecting the number of complete BUSCOs (<1% reduction). A final round of Medaka polishing improved all metrics (<1%).

De novo Mitochondrial Genome Assembly and Numt Analysis

We extracted reads containing mitogenomic sequence from the raw Nanopore data using published CO1 sequence. As there was an abundance of both small and extremely large reads containing mitonuclear sequence, we used reads that were between 16–18kb to assemble the mitochondrial genome. We assumed the mitogenome was within this range, as this range had the largest number of sequences and is the typical size for metazoan mitogenomes. Similar to that of the nuclear genome, we used Canu [24] to assemble the mitochondrial genome, followed by polishing with Medaka. The final mitogenome is 16,999 bp and contains common genes found within the mitogenomes of eukaryotes (Figure 3).

To isolate nuclear contigs that contain mitogenomic sequence, we used Geneious's annotation feature to search the final draft assembly's 3,399 contigs for mitogenomic sequence with a 25% similarity or greater with the 13 coding genes or 2 rRNAs. We found 1009 numts (989 coding sequences and 20 rRNAs) within 222 contigs, totalling 293,992 bp. One contig (Contig 171) contains a full-length mitogenomic insertion, however while the contig is verifiable using raw long reads, it is clear additional sequence was inserted during assembly that cannot be found in the long read data. Assembly data from before purging shows an additional full numt insertion that could be fully confirmed using raw long reads (Figure 3(B)). These examples demonstrate the difficulty in balancing the removal of extraneous sequence while retaining important information.

We also searched the raw Nanopore data for reads >50 kbp that contained a 50 bp match to any portion of the mitogenome. We found 118 long reads containing mitogenomic sequence, some of extreme length and 9 with complete mitogenomes (Figure 4). However, these complete numt reads are not incorporated into the final assembly.

Interestingly, the similarity of the numt coding genes and rRNAs from raw reads that map to contigs is lower than that of the contig itself. Raw reads that are ostensibly actual mitogenomic reads have typically 95% similarity or higher (less than 100% due to sequencing error) when compared to the polished mitogenome, while raw numt reads have typically 80–90% similarity. It is likely that genomic contigs are being corrected with fragmented mitochondrial reads, a potential problem for accurate genome assembly.

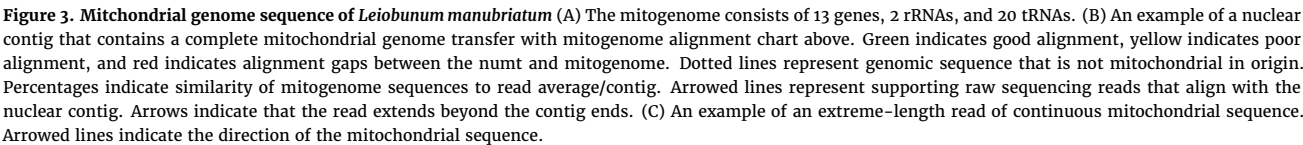
Annotation

Several datasets were used to guide annotation of the *L. manubriatum* draft genome. First, Genemark-ES (GeneMark, RRID:SCR_011930) and SNAP [29] were trained to identify protein coding genes. Second, as a transcriptome for *L. manubriatum* has not yet been generated, publicly available transcriptome RNA-seq reads from the related species *Leiobunum verrucosum* (accession num: SRR1145701) were downloaded and assembled using Trinity v2.10.0 (Trinity, RRID:SCR_013048) [30, 31]. Third, protein databases from several arthropods were downloaded from NCBI and used as references for homology prediction (SupTableX). After two rounds of training using Genemark and SNAP, we used the *L. verrucosum* transcriptome assembly and custom protein database, to guide annotation of the *L. manubriatum* genome using Maker v3.01.03 (MAKER, RRID:SCR_005309) [32, 33, 34]. The BUSCO scores for the final annotation using the arthropod gene group against predicted transcripts was 92.4% and predicted proteins was 92.3%. Furthermore, the mean AED score from Maker was 0.32, which suggests a well annotated genome.

Discussion

Genome Size

We verified the size of the *L. manubriatum* genome using Illumina HiSeq short-read sequencing data in GenomeScope [35]. Our genome size estimate is somewhat smaller than the only other publicly available nuclear genome resource for Opiliones [21], which estimates a haploid count of ~500 Mbp. Spider genomes average ~2.5 Gbp, but have a broad range from 0.74 – 5.7 Gbp [36]. Garb et al. (2018) [37] have noted a need for the resolution of additional arachnid genomes in order to answer evolutionary questions about gene duplication and its role in arachnid functional diversity. Indeed, the assembly of *P. opilio* [21] lacks whole genome



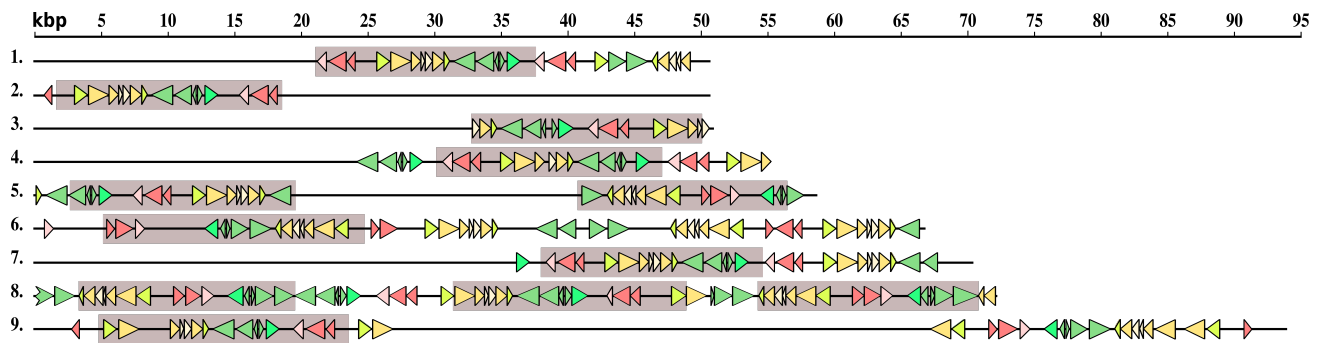


Figure 4. Raw Nanopore reads >50kb that contain a full-length mitogenome. Grey boxes indicate continuous mitogenomic sequence that includes, in order, all major genes and both rRNAs.

Nuclearized Mitochondrial Genes

We found evidence of numerous transfers of mitochondrial DNA into the nuclear genome of *L. manubriatum*. This finding is not uncommon for multi-cellular eukaryotes, which vary in numt abundance based on transfer frequency and the efficiency of nuclear gene purge [38]. However, our finding of complete mitochondrial genomes with limited interspersed nuclear sequence appears to be entirely undocumented for any arthropods (but see [10] for a mammalian example). These large blocks invite ongoing research as to the mechanism of mitochondrial gene transfer, as well as the potential and implications for functionality of these genes, which we discuss here.

Numt creation: more common in parthenogens?

While numts are common in the eukaryotic genome, little research has focused on the mechanisms responsible for their initial transfer, nor on the frequency of these transfers. Notably, chloroplastic DNA is rarely found in the nuclear genomes of plants, suggesting that organelle type may be potentially significant in the evolution of genome nuclearization. The typically small reported size of numts further suggests nuclearization is a rare event followed by generations of recombination that serve to further fragment mitochondrial genes transferred to the nuclear genome [38]. However, we posit asexually-reproducing organisms, like *L. manubriatum*, are potentially more likely to have genomes with many large numts. This is because facultative parthenogenesis, as hypothesized to occur in *L. manubriatum*, relies on meiotic errors such as nondisjunction to develop. These errors may create the germ line instability necessary to disrupt cytoplasmic separation and pull mitochondrial genes into the reforming nuclear envelope. Alternatively, organelles may incorrectly segregate to polar bodies formed during oogenesis, and later be reintroduced to the oocyte in asexual syngamy. Parthenogens are known to have larger genomes than closely related sexual species, but this is due to multiple reasons. With fewer opportunities to clear so-called "junk" DNA through outcrossing, parthenogenetic taxa accumulate transposable elements and extreme nonsynonymous mutations at higher rates than sexual species [39, 40]. Following enablement of parthenogenetic reproduction, the genomes of parthenogens frequently double due to the same meiotic errors enabling the reproductive mode itself. Thus, mitochondrial nuclearization is probably an additional contributor to the larger size of parthenogenetic genomes.

Numt maintenance: could numts be beneficial to fitness?

The large numts that we identified in the *L. manubriatum* nuclear genome were in some cases indistinguishable from the actual mitochondrial genome. This could be due to the recentness of the genomic transfers, with insufficient time in the lineage we sampled to break down the sequence of the numts via mutations and recombinatory events. However, the potential that these numts have been selectively maintained in the *L. manubriatum* genome, and even potentially transcription-active, opens a score of possibilities for novel genomic evolution. What would be the evolutionary benefit of numts within the nuclear genome? Answers to this question are dependent upon the direction and content of the transfer. Recent studies on human mitochondrial haplogroups have identified large numts whose presence resembles biparental transmission of mitochondrial DNA [41]. This is significant because, aside from a few rare cases [42, 43], mitochondria are nearly entirely maternally transmitted in animals. With an occasional influx of mitochondrial DNA entering the nucleus due to meiotic instability, the potential for the creation of a rescue reservoir of functional mitochondrial genes is formed. Such a reservoir would be extremely beneficial for obligate or facultatively parthenogenetic organisms, which are more likely to accumulate deleterious mutations than sexual species. This mechanism could additionally enable paternal transmission of mitochondrial genes. If these genes are functional, paternally-derived numts could furthermore provide genetic rescue specifically in facultative parthenogenetic species like *L. manubriatum*, which experience at least infrequent sexual reproduction.

Practical Concerns for Genome Scaffolding

Mitochondrial sequence is commonly found in the nuclear genome [44], and we have shown that in some cases these sequences may be indistinguishable from the mitochondrial source. This impacts the function of programs such as GenomeScope [35], which excludes high copy number genes from genome size estimates via kmer coverage limits. However, genome scientists may rarely examine numts, and their presence tends to be treated more as a nuisance than as a source of evolutionary information [45]. In the age of long-read sequencing, we propose that some review of the raw reads from mitochondrial sequences is justified, particularly as the abundance of mitochondria ensures that reads from numts with internal nuclear sequence, or many mutations, will be comparatively few and therefore possible to isolate and review by hand, as we have done here (Figure 3). The analysis of

fully scaffolded long-read sequences must include identification of the numts incorporated within them and separation of true mitochondrial sequence in order to identify reproductive mode or meiotic instability. The numts recovered may differ in the recentness of their transfer, their size, and their maintenance of expression. This last factor can be impacted by the location of a numt within the nuclear genome; therefore, we primarily discuss concerns with numt detection here.

Numts that have been recently formed are more likely to be complete copies of the mitochondrion, as they have not yet been impacted by recombination or mutation. This sequence may be more likely to be expressed, as well. This means that the numt will share a high percentage of sequence similarity with the mitochondrial genome. Similarly, numts that are large and/or complete may also be improperly corrected by the mitochondrial genome during scaffolding influence because of their similarities to the source. Reducing genome size to match that of external predictions may also lead to the removal of true numt sequence, as they are often tagged as repetitive or collapsed, as demonstrated here.

Nuclear assembly with large or very complete numts should first filter by percent identity of sequenced reads to the mitochondrial genome. If the assembly goals do not include analysis of numts, a cutoff value can be employed to remove all high copy reads from mitochondrial assemblies to ensure that the mitochondrial genome does not influence the nuclear consensus sequence by erroneously correcting any numts. If, however, there is interest in studying the numts, filtration to remove mitochondrial reads with a length equal to or shorter than the mitochondrial genome should be performed to ensure that numts are not corrected. Reads containing internal sequence that does not map to the mitochondrion could later be isolated and returned to the pool of fragments for assembly. This procedure would therefore preserve numts for downstream study.

Data Availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/NCBI GenBank under the project number PRJNA814647.

Declarations

List of abbreviations

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations should be provided in alphabetical order.

BUSCO = Benchmark Universal Single Copy Ortholog

Competing Interests

The author(s) declare that they have no competing interests.

Ethics Approval and Consent to Participate

Not applicable.

Consent to Publication

Not applicable.

Funding

This research was supported in part by a UMBC START grant to M.B. and in part by a National Science Foundation IOS grant to both authors (M.B.: 2113665; S.S.: 2113666). The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data, nor in the decision to publish.

Author's Contributions

M.B conceived of and planned the experiments; M.B and S.S. collected specimens and extracted DNA; S.S. conducted sequencing runs and assembled the genome; M.B. and S.S co-wrote the manuscript.

Acknowledgements

We thank Dr. Nobuo Tsurusaki for supporting our collection efforts in Japan, including making housing arrangements, driving to the collecting sites, and allowing generous use of his laboratory. Preliminary analyses were conducted with the UMBC High Performance Computing Facility.

References

- Simon JC, Delmotte F, Rispe C, Crease T. Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society* 2003;79(1):151–163.
- Otto SP, Whitton J. Polyloid incidence and evolution. *Annual Review of Genetics* 2000;34(1):401–437.
- Neiman M, Kay AD, Krist AC. Can resource costs of polyploidy provide an advantage to sex? *Heredity* 2013;110(2):152–159.
- Dufresne F. The history of the *Daphnia pulex* complex. In: Christoph Held, Stefan Koenemann CS, editor. *Phylogeography and Population Genetics in Crustacea*, vol. 19 CRC Press; 2011.p. 217–232.
- Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G. Wide variation in ploidy level and genome size in a New Zealand freshwater snail with coexisting sexual and asexual lineages. *Evolution* 2011;65(11):3202–3216.
- Burke NW, Crean AJ, Bonduriansky R. The role of sexual conflict in the evolution of facultative parthenogenesis: A study on the spiny leaf stick insect. *Animal Behaviour* 2015;101:117–127.
- Tsurusaki N. Parthenogenesis and Geographic Variation of Sex Ratio in Two Species of *Leiobunum* (Arachnida, Opiliones). *Zoological Science* 1986;3:517–532.
- Burns M, Hedin M, Tsurusaki N. Population genomics and geographical parthenogenesis in Japanese harvestmen (Opiliones, Sclerosomatidae). *Ecology Evolution* 2017;8(1):36–52.
- Leister D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends in Genetics* 2005;21(12):655–663.
- Schmitz J, Noll A, Raabe CA, Churakov G, Voss R, Kieffmann M, et al. Genome sequence of the basal haplorrhine primate *Tarsius syrichta* reveals unusual insertions. *Nature Communications* 2016;7(1):1–11.
- Matsui A, Rakotondrapary F, Munekika I, Hasegawa M, Horai S. Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs. *Gene* 2009;441(1–2):53–66.
- Mourier T, Hansen AJ, Willerslev E, Arctander P. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Molecular Biology and Evolution* 2001;18(9):1833–1837.
- Pamilo P, Viljakainen L, Vihavainen A. Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution* 2007;24(6):1340–1346.
- Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: Mt-Genome organization and evolution in nonbilaterian lineages. *Genome Biology and Evolution* 2016;8(9):2896–2913.
- Ko BJ, Chul L, Kim J, Rhie A, Yoo DA, Howe K, et al. Widespread false gene gains caused by duplication errors in genome assemblies. *Genome Biology* 2020;23(205):1–26.
- Prodanov T, Bansal V. Sensitive alignment using paralogous sequence variants improves long-read mapping and variant calling in segmental duplications. *Nucleic Acids Research* 2020;48(19):e114.
- Zhang DX, Hewitt GM. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* 1996;11(6):247–251.
- Torres L, Welch AJ, Zanchetta C, Chesser RT, Manno M, Donnadieu C, et al. Evidence for a duplicated mitochondrial region in Audubon's shearwater based on MinION sequencing. *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis* 2019;30(2):256–263.
- Dhorne-Pollet S, Barrey E, Pollet N. A new method for long-read sequencing of animal mitochondrial genomes: application to the identification of equine mitochondrial DNA variants. *BMC Genomics* 2020;21(785):1–15.
- Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, et al. Complete vertebrate mitogenomes reveal widespread gene repeats and gene duplications. *Genome Biology* 2021;22(120):1–22.
- Gainett G, González VL, Ballesteros JA, Setton EVW, Baker CM, Barolo Gargiulo L, et al. The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. *Proceedings of the Royal Society B: Biological Sciences* 2021;288(1956):2021.01.11.426205. <https://doi.org/10.1101/2021.01.11.426205>.
- Sperling AL, Fabian DK, Garrison E, Glover DM. A genetic basis for facultative parthenogenesis in *Drosophila*. *Current Biology* 2023;33(17):3545–3560.
- Robinson JA, Bowie RC, Dudchenko O, Aiden EL, Hendrickson SL, Steiner CC, et al. Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology* 2021;31(13):2939–2946.
- Koren S, Walenz BP, Berlin K, Miller JR, M PA. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 2017;27:722–736.
- Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 2018;19(1):460.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31(19):3210–3212.
- Sheffer MM, Hoppe A, Krehenwinkel H, Uhl G, Kuss AW, Jensen L, et al. Chromosome-level reference genome of the European wasp spider *Argiope bruennichi*: a resource for studies on range expansion and evolutionary adaptation. *GigaScience* 2021;10:1–12.
- Sánchez-Herrero JF, Frías-López C, Escuer P, Hinojosa-Alvarez S, Arnedo MA, Sánchez-Gracia A, et al. The draft genome sequence of the spider *Dysdera silvatica* (Araneae, Dysderidae): A valuable resource for functional and evolutionary genomic studies in chelicerates. *GigaScience* 2019;8(8):1–9.
- Li S, Ma L, Li H, Vang S, Hu Y, Boland L, et al. SNAP: an integrated SNP annotation platform. *Nucleic Acids Research* 2007;35:D707–D710.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* 2011;29(7):644–52.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 2013;8(8):1494–512.
- Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 2008;18(1):188–196.
- Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome

- projects. *BMC Bioinformatics* 2011;12:491.
34. Campbell MS, Holt C, Moore B, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Current Protocols in Bioinformatics* 2014;48:4.11.1–4.11.39.
 35. Vurtture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads 2017;33(14):2202–2204.
 36. Gregory TR, Shorthouse DP. Genome Sizes of Spiders. *Journal of Heredity* 2003 07;94(4):285–290.
 37. Garb JE, Sharma PP, Ayoub NA, Recent progress and prospects for advancing arachnid genomics; 2018.
 38. Richly E, Leister D. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Molecular Biology and Evolution* 2004;21(10):1972–1980.
 39. Sharbrough J, Luse M, Boore JL, Logsdon Jr JM, Neiman M. Radical amino acid mutations persist longer in the absence of sex. *Evolution* 2018;72(4):808–824.
 40. McElroy K, Muller S, Lamatch DK, Bankers L, Fields PD, Jalinsky JR, et al. Asexuality associated with marked genomic expansion of tandemly repeated rRNA and histone genes. *Molecular Biology and Evolution* 2021;38(9):3581–3592.
 41. Bai R, Cui H, Devaney JM, Allis KM, Balog AM, Liu X, et al. Interference of nuclear mitochondrial DNA segments in mitochondrial DNA testing resembles biparental transmission of mitochondrial DNA in humans. *Genetics in Medicine* 2021;23:1514–1521.
 42. Breton S, Stewart DT. Atypical mitochondrial inheritance patterns in eukaryotes. *Genome* 2015;58(10):423–431.
 43. Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, et al. Biparental inheritance of mitochondrial DNA in humans. *PNAS* 2018;115(51):13039–13044.
 44. Hazkani-Covo E, Zeller RM, Martin W. Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *PLOS Genetics* 2010;6(2):1–11. <https://doi.org/10.1371/journal.pgen.1000834>.
 45. Graham NR, Gillespie RG, Krehenwinkel H. Towards eradicating the nuisance of numts and noise in molecular biodiversity assessment. *Molecular Ecology Resources* 2021;21(6):1755–1758.