GBE

Genome of Facultatively Parthenogenetic Harvester (Opiliones) Indicates Frequent Mitonuclear Sequence Transfer and Novel Full-Length Insertions

Sarah D. Stellwagen [1,*], Mercedes Burns [2]

[1]Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, USA
[2]Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, USA

*Corresponding author: E-mail: stellwagen@charlotte.edu.

Accepted: April 24, 2026

Abstract

Facultative parthenogenesis and intrapopulation mixed ploidy (where individuals with differing ploidy exist within the same area) are rare in animals. However, these unique characteristics allow opportunities to investigate the relationship between sexual modality and genome structure. We have completed a genome assembly of the Japanese harvester ("daddy-longlegs") *Leiobunum manubriatum*, a species that reproduces sexually and asexually and has mixed diploid and tetraploid populations. 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 Arthropoda (complete: 97.7%). We also 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 nine reads within the raw sequencing data that were collapsed during the assembly process. The genome of the *L. manubriatum* diploid race is a valuable resource not only for opilionid 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.

Significance

We sequenced the genome of the *Leiobunum manubriatum* diploid race and found evidence of several full-length nuclearized mitochondrial sequences (numts) within the genome. We discuss the potential correlation between asexual reproductive modes and numt generation and provide practical suggestions for handling large numts in long-read sequencing projects. This work represents the second genome from the order Opiliones to be sequenced and the first genome of wild-derived facultatively parthenogenetic species to be published.

Introduction

Sexual reproduction is the most common reproductive mechanism for multicellular eukaryotes. However, approximately one out of every 1000 multicellular eukaryotes exhibits parthenogenesis (Simon et al. 2003), and the genomes of these species are expected to display particular associated features. Asexual reproduction is often associated with polyploidy in plant and animal systems (Otto and Whitton 2000; Neiman et al. 2013) because the mechanisms that enable asexual reproduction, such as chromosomal nondisjunction or interspecific hybridization, also cause gene duplications (Barley et al. 2021). For example, the *Daphnia pulex* complex of water fleas has cyclically sexual/asexual populations, in addition to polyploid, obligately asexual lineages (Dufresne 2011). Populations of New Zealand freshwater snail *Potamopyrgus antipodarum*, which include sexual and asexual lineages, have variations in ploidy and genome size (Neiman et al. 2011) that also correlate with reproductive mode.

Nuclearized mitochondrial sequences (numts) are another genomic feature that could covary with parthenogenesis. Numts are regions of the nuclear genome originating from the mitochondrial genome (Leister 2005). They are common in eukaryotes, but the mechanisms underlying their initial transfer remain unclear. Numts tend to be fragmented and, to our knowledge, only two full-length mitogenome insertions to the nuclear genome have been reported (tarsiers: Schmitz et al. 2016; *Arabidopsis thaliana*: Fields et al. 2022). The largest human numt covers approximately 90% of the mitochondrial genome (Mourier et al. 2001), and large numts of other mammals have been identified (Wang et al. 2015; Dayama et al. 2020; Bolner et al. 2024). In arthropods, honeybees have the highest percentage of reported numts, though these sequences are relatively short (Pamilo et al. 2007). There has been no research examining numts in parthenogenetic animals, potentially owing to the size and complexity of their genomes. Mitochondrial genomes average ~17 kbp (Lavrov and Pett 2016), and duplications can confuse assembly efforts when aligning sequences obtained through short-read technologies (Ko et al. 2020; Prodanov and Bansal 2020).

Two species of nonspider arachnids in the order Opiliones (also known as "harvesters" or "daddy-longlegs") exhibit facultative parthenogenesis: *Leiobunum manubriatum* (Fig. S1) and *Leiobunum globosum*. These species are endemic to northern Japan, with *L. manubriatum*'s range extending through the Japanese Alps and overlapping with that of *L. globosum* in Aomori, Akita, and Hokkaido Prefectures. While *L. globosum* is entirely tetraploid (Tsurusaki 1986), flow cytometry has shown that some populations of *L. manubriatum* have diploid and tetraploid individuals, while others have a single cytotype (Burns et al. 2017).

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 genome is the first of a facultatively parthenogenetic species and the second for Opiliones following Gainett et al. (2021). The *L. manubriatum* nuclear genome has unusually large numt insertions and documents the first incidence of multiple full-length transfers in animals.

Results and Discussion

The final assembly of diploid *L. manubriatum* is 336 Mbp from 3,399 contigs, with an N50 of 27,489,741 bp (Table 1; NCBI Project: PRJNA814647). The scaffolding process resulted in eight (out of 12 expected; Tsurusaki 1986) complete chromosomes of the haploid genome (chromosome sizes reported in Table S1); however, telomeric sequence was not recovered. Half of the genome is represented by four contigs (L50). The polishing program Medaka (Oxford Nanopore) and heterozygous haplotype reduction tool Purge Haplotigs (Roach et al. 2018) greatly improved BUSCO completeness metrics while reducing duplications (Fig. S2). The genome recovered 97.7% complete BUSCOs (Simão et al. 2015) of the 1,066 arthropod genes (Table 1; additional BUSCOs in Table S2). These BUSCO scores are excellent

Table 1 *L. manubriatum* nuclear genome assembly statistics with BUSCO completeness scores

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	71,836,609
N50 (bp)	27,489,741
L50 (bp)	4
Protein-coding genes	24,032
BUSCO scores	
Assembly, Arthropoda	C: 97.7% [S: 92.8%, D: 4.9%], F: 1%, M: 1.3%
Annotation, Arthropoda, transcripts	C: 93.6% [S: 88.3%, D: 5.3%], F: 2.7%, M: 3.7%
Annotation, Arthropoda, proteins	C: 93.6% [S: 88.5%, D: 5.1%], F: 2.7%, M: 3.7%

C, complete [S + D]; S, single; D, duplicated; F, fragmented; M, missing.

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 (Sheffer et al. 2021). Coverage and N50 statistics for individual Nanopore runs are reported in Table S4, and the estimated genome size, using GenomeScope (Vurture et al. 2017), is approximately 306 Mbp (Fig. S4). The BUSCO scores for the final annotation using the arthropod gene group against the longest isoforms of predicted transcripts and predicted proteins were both 93.6%. Furthermore, the mean AED (annotation edit distance) score from Maker (Cantarel et al. 2008) was 0.32. An AED score of <0.5 is considered good, while scores ≤0.3 are indicative of a high-quality annotation.

The final mitogenome is 16,999 bp and contains genes common to eukaryotic mitogenomes (Fig. 1a). We found 1,009 numts (989 coding sequences and 20 rRNAs) within 222 contigs, totaling 293,992 bp, including 155 numts in the eight reported chromosomes (Fig. 1b). One contig (contig ID 171) contains a full-length mitogenomic insertion. However, while the contig is verifiable using raw long reads, additional sequence was inserted during assembly that is not

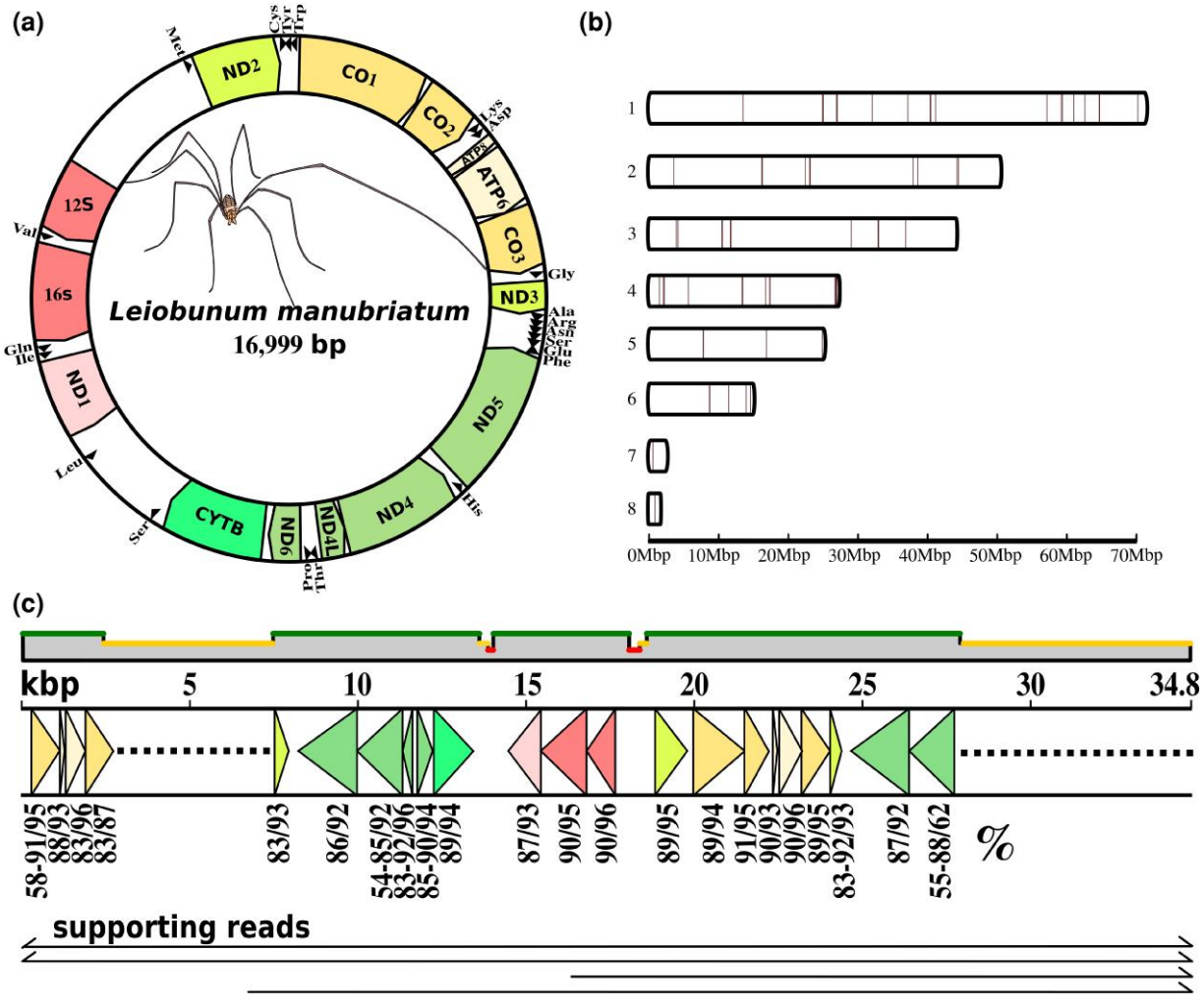

Fig. 1. Mitochondrial genome sequence of *L. manubriatum*, chromosomal locations of numts, and a nuclear contig with a complete numt. a) The mitogenome consists of 13 genes, 2 rRNAs, and 20 tRNAs. b) Locations of numts with at least 50% similarity to corresponding regions of the mitogenome within the eight completed chromosomes of the assembly. Thicker lines represent clusters of numts that cannot be individually resolved at this scale. c) Example nuclear contig (contig ID 195) that contains a complete numt with mitogenome alignment chart above. Green indicates good alignment, yellow indicates poor alignment, and red indicates alignment gaps between the numt and mitogenome (from Geneious mean pairwise identity metrics scores and visualization). Dotted lines represent genomic sequence that is not mitochondrial in origin. Percentages indicate similarity of mitogenomic genes compared to those in the contig from the polished assembly or raw reads that map to that contig (raw read/contig). Arrowed lines represent supporting raw sequencing reads that align with the nuclear contig. Arrows indicate that the read extends beyond the contig ends.

present in the long-read data. Prepurge assembly data shows an additional full numt insertion that could be fully confirmed using raw long reads (Fig. 1c). These examples demonstrate the difficulty in balancing the removal of mitochondrial sequence while retaining nuclear information.

We queried the raw Nanopore data and found 118 long reads (>50 kbp) containing mitogenomic sequence, nine with complete mitogenomes (Fig. S3). However, these complete numt reads are not incorporated into the final assembly. Interestingly, genes from the polished mitogenome typically have 95% similarity or higher (less than 100% due to sequencing error) to those in assembly contigs, while these genes in raw reads (which are ostensibly more accurate representations of transferred mitogenomic sequence) typically have 80% to 90% similarity (Fig. 1c). It is possible that genomic contigs are being corrected with fragmented mitochondrial reads, which could confuse our understanding of numt degradation processes and be a potential problem for accurate genome assembly discussed below.

We estimate the genome size of *L. manubriatum* to be approximately 336 Mbp. Our estimate is somewhat smaller than the only other publicly available nuclear genome resource for Opiliones (Gainett et al. 2021), which estimates a haploid count of ∼500 Mbp. Notably, this assembly lacks whole genome duplications found in other arachnid lineages, such as spiders (Gregory and Shorthouse 2003). Ongoing genome evolution research in arachnids will benefit from improved assemblies that incorporate long reads (Garb et al. 2018), such as in this work.

We found evidence of numerous transfers of mitochondrial DNA into the nuclear genome of *L. manubriatum*. This is typical for multicellular eukaryotes, which vary in numt abundance based on transfer frequency and the efficiency of gene turnover (Richly and Leister 2004). Our finding of complete numts appears to be entirely undocumented for any arthropods, although this may be because numts are rarely examined and/or cannot be conclusively identified due to the sequencing and assembly methods used (Hlaing et al. 2009). Numts tend to be treated more as a nuisance than as a source of evolutionary information (Graham et al. 2021). We posit that asexually reproducing organisms, such as *L. manubriatum*, are potentially more likely to have genomes with many large numts. This is because thelytokous parthenogenesis, as observed in *L. manubriatum*, can develop from meiotic errors, such as chromosomal nondisjunction (Yagound et al. 2020). These errors may create the germ line instability necessary to disrupt cytoplasmic separation and pull mitochondrial genes into the reforming nuclear envelope.

We have shown that numt sequences may be indistinguishable from the mitochondrial source. This impacts barcoding analyses (Song et al. 2008) and the function of programs, such as GenomeScope, which excludes high copy number genes from genome size estimates via kmer coverage limits. Numts that are large and/or complete may be improperly corrected by the mitochondrial genome during scaffolding 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. We therefore propose that 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 (Fig. S3). An alternative strategy comparing genomic and cDNA-derived mitochondrial amplicons may also be appropriate (Lopez et al. 2021).

Nuclear assembly with large or very complete numts should first be filtered 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 there is interest in studying 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 sequences that do 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.

Methods

Sample Collection and Extraction

Adult female *L. manubriatum* specimens were collected from diploid-only populations in forest around Hirayu Campground on July 11, 2014, and August 3, 2019, and Shōmyo Falls visiting area on July 11, 2014, and August 3 to 4, 2019 (Fig. S1). The ploidy of these populations was determined earlier via karyotyping (Tsurusaki 1986) and checked with flow cytometry (Burns, unpublished data). Permits are not required for arthropod collection on public lands in Japan; thus, all tissues were collected in adherence to the Nagoya protocol. Specimens collected in 2014 were stored in 100% ethanol. Specimens collected in 2019 were

immediately transported live to Tottori University, Japan, for DNA extraction. To reduce contamination from gut flora and parasites, the gut of each specimen was removed. High molecular weight DNA was extracted from the remaining tissue of each specimen using the MasterPure Complete DNA Purification Kit (cat. no. MC89010) and transported to the University of Maryland, Baltimore County, for further processing.

Nanopore Sequencing

The DNA from 29 diploid specimens was used for sequencing. Due to their size, the amount of DNA needed for high molecular weight DNA isolation, and the required sequencing library load, samples were combined. Extracted DNA was pooled to obtain 10 ug 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 cleaned using Agencourt AMPure XP beads (cat. no. A63881) and eluted overnight to several days in water. Purified 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 fasta 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) (Koren et al. 2017) with default parameters. Preserved tissue from specimens with gut removed was shipped to Dovetail Genomics, and the raw draft assembly was then further scaffolded by Dovetail HiRise. This draft assembly was then further polished using Nanopore's Medaka v1.0.1 program. Purge Haplotigs v1.1.1 (Roach et al. 2018) was then used on the polished assembly to remove heterozygous haplotype contigs that were assembled separately with a = 50. We used GenomeScope and Illumina HiSeq short-read sequencing data obtained from whole genome shotgun sequencing of a single individual to confirm genome size (Vurture et al. 2017). Consistent with other arthropod assemblies (Francois et al. 2020), contaminants identified by the NCBI FCS-GX screen (Astashyn et al. 2024) in the *L. manubriatum* assembly included microsporidians, spirochaetes, proteobacteria and associated viruses. Approximately 2.3 Mbp were removed from the initial *L. manubriatum* assembly, with alpha-proteobacteria accounting for the largest proportion of foreign DNA (26%).

De novo Mitochondrial Genome Assembly and Numt Analysis

Using the Geneious custom BLAST feature to search the raw Nanopore data with published *L. manubriatum* COI sequence (accession number: MG201509.1) as the query, we extracted reads containing mitogenomic sequences. We used reads that were between 16 and 18 kb to assemble the mitochondrial genome. Similar to that of the nuclear genome, we used Canu (Koren et al. 2017) to assemble the mitochondrial genome, followed by polishing with Medaka.

To isolate nuclear contigs with 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 two rRNAs. We used a MATLAB script (see Supplementary Material) to search the raw Nanopore data for reads > 50 kbp and containing a 50 bp exact match to any portion of the mitogenome.

Annotation

Several datasets were used to guide annotation of the *L. manubriatum* genome. First, we trained GeneMark-ES (GeneMark, RRID:SCR_011930) and SNAP (Li et al. 2007) to identify protein-coding genes. Second, as a transcriptome for *L. manubriatum* has not yet been generated, we downloaded and assembled publicly available transcriptome RNA-seq reads from *Leiobunum verrucosum* (accession number: SRR1145701) using Trinity v2.10.0 (Trinity, RRID:SCR_013048) (Grabherr et al. 2011; Haas et al. 2013). Third, we referenced protein databases from several arthropods from NCBI for homology prediction (Table S3). After two iterations with 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) (Holt and Yandell 2011; Campbell et al. 2014).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Dr. Nobuo Tsurusaki for supporting our work in Japan, including arranging housing and transportation and allowing generous use of his laboratory. Dr. John Wang and an anonymous reviewer provided helpful comments that improved the manuscript. Preliminary

analyses were conducted with the UMBC High Performance Computing Facility.

Author Contributions

M.B. 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.

Funding

This research was supported by a University of Maryland Baltimore County START (to M.B.) and the National Science Foundation (2113666 to S.S. and 2113665 and 2237684 to M.B.). The funding bodies had no role in the design of the study and collection, analysis, or interpretation of data, nor in the decision to publish.

Data Availability

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

Literature Cited

Astashyn A, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol. 2024:25:60. https://doi.org/10.1186/s13059-024-03198-7.

Barley AJ, Reeder TW, Nieto-Montes de Oca A, Cole CJ, Thomson RC. A new diploid parthenogenetic whiptail lizard from Sonora, Mexico, is the "missing link" in the evolutionary transition to polyploidy. Am Nat. 2021:198:295–309. https://doi.org/10.1086/715056.

Bolner M, et al. A comprehensive atlas of nuclear sequences of mitochondrial origin (NUMT) inserted into the pig genome. Genet Sel Evol. 2024:56:64. https://doi.org/10.1186/s12711-024-00930-6.

Burns M, Hedin M, Tsurusaki N. Population genomics and geographical parthenogenesis in Japanese harvestmen (Opiliones, Sclerosomatidae). Ecology Evolution. 2017:8:36–52. https://doi.org/10.1002/ece3.3605.

Campbell MS, Holt C, Moore B, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Curr Protoc Bioinformatics. 2014:48:4.11.1–4.11.39. https://doi.org/10.1002/0471250953.bi0411s48.

Cantarel BL, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008:18:188–196. https://doi.org/10.1101/gr.6743907.

Dayama G, Zhou W, Prado-Martinez J, Marques-Bonet T, Mills RE. Characterization of nuclear mitochondrial insertions in the whole genomes of primates. NAR Genom Bioinform. 2020:2:lqaa089. https://doi.org/10.1093/nargab/lqaa089.

Dufresne F. The history of the daphnia pulex complex. In: Christoph Held, Stefan Koenemann CS, editors. Phylogeography and population genetics in Crustacea. Vol. 19: CRC Press; 2011. p. 217–232.

Fields PD, et al. Complete sequence of a 641-kb insertion of mitochondrial DNA in the Arabidopsis thaliana nuclear genome. Genome Biol Evol. 2022:14:evac059. https://doi.org/10.1093/gbe/evac059.

Francois CM, Durand F, Figuet E, Galtier N. Prevalence and implications of contamination in public genomic resources: a case study of 43 reference arthropod assemblies. G3 Genes Genomes Genetics. 2020:10:721–730. https://doi.org/10.1534/g3.119.400758.

Gainett G, et al. The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. Proc R Soc Lond B Biol Sci. 2021:288:20211168. https://doi.org/10.1098/rspb.2021.1168.

Garb JE, Sharma PP, Ayoub NA. Recent progress and prospects for advancing arachnid genomics. Curr Opin Insect Sci. 2018:25:51–57. https://doi.org/10.1016/j.cois.2017.11.005.

Grabherr MG, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 2011:29:644–652. https://doi.org/10.1038/nbt.1883.

Graham NR, Gillespie RG, Krehenwinkel H. Towards eradicating the nuisance of numts and noise in molecular biodiversity assessment. Mol Ecol Resour. 2021:21:1755–1758. https://doi.org/10.1111/1755-0998.13414.

Gregory TR, Shorthouse DP. Genome sizes of spiders. J Hered. 2003:94:285–290. https://doi.org/10.1093/jhered/esg070.

Haas BJ, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013:8:1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hlaing T, et al. Mitochondrial pseudogenes in the nuclear genome of Aedes aegypti mosquitoes: implications for past and future population genetic studies. BMC Genet. 2009:10:11. https://doi.org/10.1186/1471-2156-10-11.

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011:12:491. https://doi.org/10.1186/1471-2105-12-491.

Ko BJ, et al. Widespread false gene gains caused by duplication errors in genome assemblies. Genome Biol. 2020:23:205. https://doi.org/10.1186/s13059-022-02764-1.

Koren S, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017:27:722–736. https://doi.org/10.1101/gr.215087.116.

Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: Mt-Genome organization and evolution in nonbilaterian lineages. Genome Biol Evol. 2016:8:2896–2913. https://doi.org/10.1093/gbe/evw195.

Leister D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet. 2005:21:655–663. https://doi.org/10.1016/j.tig.2005.09.004.

Li S, et al. SNAP: an integrated SNP annotation platform. Nucleic Acids Res. 2007:35:D707–D710. https://doi.org/10.1093/nar/gkl969.

Lopez MLD, et al. Using metatranscriptomics to estimate the diversity and composition of zooplankton communities. Mol Ecol Resour. 2021:22:638–652. https://doi.org/10.1111/1755-0998.13506.

Mourier T, Hansen AJ, Willerslev E, Arctander P. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol Biol Evol. 2001:18:1833–1837. https://doi.org/10.1093/oxfordjournals.molbev.a003971.

Neiman M, Kay AD, Krist AC. Can resource costs of polyploidy provide an advantage to sex? Heredity (Edinb). 2013:110:152–159. https://doi.org/10.1038/hdy.2012.78.

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. 2011:65:3202–3216. https://doi.org/10.1111/j.1558-5646.2011.01360.x.

Otto SP, Whitton J. Polyploid incidence and evolution. Annu Rev Genet. 2000:34:401–437. https://doi.org/10.1146/annurev.genet.34.1.401.

Pamilo P, Viljakainen L, Vihavainen A. Exceptionally high density of NUMTs in the honeybee genome. Mol Biol Evol. 2007:24:1340–1346. https://doi.org/10.1093/molbev/msm055.

Prodanov T, Bansal V. Sensitive alignment using paralogous sequence variants improves long-read mapping and variant calling in segmental duplications. Nucleic Acids Res. 2020:48:e114. https://doi.org/10.1093/nar/gkaa829.

Richly E, Leister D. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Mol Biol Evol. 2004:21:1972–1980. https://doi.org/10.1093/molbev/msh210.

Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018:19:460. https://doi.org/10.1186/s12859-018-2485-7.

Schmitz J, et al. Genome sequence of the basal haplorrhine primate *Tarsius syrichta* reveals unusual insertions. Nat Commun. 2016:7:12997. https://doi.org/10.1038/ncomms12997.

Sheffer MM, 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:giaa148. https://doi.org/10.1093/gigascience/giaa148.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Simon JC, Delmotte F, Rispe C, Crease T. Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biol J Linn Soc Lond. 2003:79:151–163. https://doi.org/10.1046/j.1095-8312.2003.00175.x.

Song H, Buhay JE, Whiting MF, Crandall KA. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci U S A. 2008:105:D707–D710. https://doi.org/10.1073/pnas.0803076105.

Tsurusaki N. Parthenogenesis and geographic variation of sex ratio in two species of Leiobunum (Arachnida, Opiliones). Zoolog Sci. 1986:3:517–532. https://doi.org/10.34425/zs000259.

Vurture GW, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017:33:2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

Wang B., et al. Full-length Numt analysis provides evidence for hybridization between the Asian colobine genera Trachypithecus and Semnopithecus. Am J Primatol 2015:77:901–910. https://doi.org/10.1002/ajp.22419

Yagound B, et al. A single gene causes thelytokous parthenogenesis, the defining feature of the cape honeybee *Apis mellifera capensis*. Curr Biol. 2020:30:2248–2259. https://doi.org/10.1016/j.cub.2020.04.033.

Associate editor: John Wang