diversity

MDPI

Article

DIY Mitogenome of Your Favorite Animal: Benchmarking Mitochondrial Genome Assembly and Annotation Tools for Non-Model Species Using Lake Baikal Amphipods

Andrei Mutin [1], Iuliia Ozerova [2], Maxim Timofeyev [1] and Polina Drozdova [1,*]

[1] Institute of Lake Baikal Biology, Irkutsk State University, Irkutsk 664025, Russia; andreimutin97@gmail.com (A.M.); m.a.timofeyev@gmail.com (M.T.)
[2] Institute for Medical Biometry and Bioinformatics, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany; iuliia.ozerova@hhu.de
* Correspondence: drozdovapb@gmail.com

Abstract

The mitochondrial genome is arguably the most frequently studied part of the eukaryotic genome and the source of the major DNA barcoding marker for animals: the 5′ region of the cytochrome c oxidase subunit I gene (*cox1*, COI, or CO1). The majority of animal mitochondrial genomes are small circular DNA molecules (~16 kbp) with high copy numbers and should therefore assemble easily. However, assembling mitochondrial genomes from short reads still poses challenges. Our experience assembling mitochondrial genomes of amphipods (Crustacea: Amphipoda) and our review of the literature showed that, in most cases, automatic pipelines failed, and that many studies used highly computationally intensive pipelines that require high-performance computers. Here, we compared ten assemblers and six mitochondrial genome annotators, using raw sequencing data from several Baikal amphipod species with nuclear genomes varying in size and complexity, based on automation, ease of use for non-bioinformaticians, and low computational resource requirements. We conclude that there is no single best solution. While MITObim and MitoFinder produced the best reference-based and de novo assemblies, respectively, from DNA data, MitoZ returned the most consistent results for RNA sequencing data. Among annotation pipelines, MitoFinder was the closest to transferring annotation from a reference and MITOS2 provided the most reliable de novo annotation. We also recommend a procedure for assembling and annotating a new mitochondrial genome that worked best on amphipod genome sequencing data. The reproducible code is available at GitHub.

Keywords: lcWGS; genome skimming; mitogenome assembly; mitogenome annotation; non-model species

Academic Editor: Luc Legal

Received: 16 October 2026
Revised: 13 November 2026
Accepted: 23 November 2026
Published: 27 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Mitochondrial Genome Assembly and Annotation: Choices and Challenges

The mitochondrial genome is arguably the most frequently studied part of the eukaryotic genome [1] and the source of the major DNA barcoding sequence for animals: the 5′ region of the cytochrome c oxidase subunit I gene (*cox1*, COI, or CO1) [2,3]. Important features that have contributed to the success of mitochondrial DNA as a genetic

marker include high copy number, which facilitates amplification, and limited recombination, mostly uniparental inheritance and low frequency of long indels [4,5]. Typical invertebrate mitochondrial genomes are small (15–20 kbp), circular molecules containing 37 genes: 13 protein-coding, 2 rRNA, and 22 tRNA genes. However, these patterns have rare but notable exceptions: exceptionally long (up to 80 kbp) [6] and/or linear mitogenomes [7], as well as multiple cases of divergence from the standard gene set [8], have been described. The most unusual taxonomic groups in this regard are basal groups of Metazoa, namely Cnidaria, Porifera, Ctenophora, and Placozoa [8,9], as well as Mollusca [10] and Isopoda [7].

Even though complete animal mitochondrial genomes are incomparably smaller, less complex, and therefore much easier to assemble than complete nuclear genomes, this process still requires many adjustments and choices along the way. The first choice facing anyone starting a mitochondrial genome project is between sample preparation and sequencing technologies: (1) long-range PCR; (2) next-generation short-read sequencing (NGS); or (3) third-generation long-read sequencing (TGS). Long-range PCR produces sequences that are easy to assemble but require considerable wet-lab optimization [11]. NGS data are currently the most prevalent in public databases, but they present some challenges in terms of assembly (see below). Techniques employing direct long-read sequencing of mitochondrial DNA, both with and without enrichment, have been shown to greatly facilitate assembly in terms of required time and effort [12–15]. However, short-read data still prevail. Aside from low-coverage whole-genome sequencing projects, mitochondrial reads are useful by-products of target sequence enrichment methods [16]. In addition, short reads are essential for museomics, which deals with degraded DNA samples and is generally incompatible with long-read technologies [17].

Another question that might arise at the inception of any mitochondrial genome project is whether to sequence genomic DNA or complementary DNA (RNA-seq). Mitochondrial reads also accumulate in RNA-seq experiments and can be used to retrieve mitochondrial genes and, in some cases, even reconstruct complete mitogenomes [18–20]. However, recovering tRNA genes and the control region from RNA-seq data is problematic due to the very uneven read depth, which is low in these regions [20]. While RNA-seq requires higher sample handling effort, it offers a competitive advantage by providing an easier and cheaper route to genome-wide datasets for functional analysis and phylogenomics than low-coverage whole-genome sequencing (lcWGS) for species with large, complex genomes [21,22]. Long RNA-seq reads can also be used to reconstruct complete mitochondrial genomes due to pervasive transcription [23].

In response to the need to assemble short-read data, a number of specialized tools for mitochondrial (or both mitochondrial and plastid) genome assembly from DNA and RNA sequencing have been suggested and are being developed [16,24–35]. They differ in the presence and implementation of a read-filtering step, in the tool used for assembly, and in their reliance on a reference sequence, which can serve as a bait for read retrieval, a query for finding mitochondrial contigs, or not be needed at all (see Section 2). Several pipelines that reuse other tools and streamline data retrieval from public databases have also been published [36,37]. For almost all applications, the assembled mitochondrial genome needs annotation; in some cases, annotation is used to assess the quality of the assembly and is included in a common pipeline. Annotation of the mitochondrial genome comprises several independent processes, namely identifying borders and assigning function for tRNA, rRNA, and protein-coding genes; some tools can also annotate the control region of the mtDNA. Protein-coding and rRNA genes are mostly annotated by similarity using specific models, since the mitochondrial genetic code varies from the standard code and rRNA structures differ from the nuclear machinery. Annotation of tRNA genes can

also be complicated: while cloverleaf structures are well defined, truncated shapes of mitochondrial tRNAs are not rare in Metazoa, and not all currently existing annotation methods take them into account [38]. Additionally, tRNA gene overlaps, which are resolved in the mitochondrion during RNA editing [39], can confuse models and require careful manual annotation. It has also been shown that mitochondrial tRNAs can change identity through genomic remolding, the process of switching tRNA identity that can happen through gene duplication followed by anticodon mutation and loss of the ancestral copy [40,41]. The resulting tRNA genes have structural features of one tRNA family but function as another, which can lead to misannotation when using standard approaches.

Another challenge in mitochondrial genome annotation is the annotation of the control region (CR), which plays an essential role in DNA replication and transcription. In well-studied mammalian mitogenomes, there are two origins of replication, one for each DNA strand: the heavy-strand origin, O_H, located inside the CR, and the light-strand origin, O_L, located around a cluster of five tRNA genes [42,43]. Among invertebrates, most experimental data come from insects, in which O_H is found within the CR, but its position varies [44]. O_L is studied less, but in *Drosophila* and the ancestral arthropod gene order it is also located in the CR [45]. Moreover, it has been shown that mitochondrial DNA replication in *Drosophila* predominantly follows the strand-coupled model, which uses a single origin [43]. Finally, information on Amphipoda, which comes from computational analysis of nucleotide composition, strongly suggests high variability of the replication origin position due to multiple inversions involving this element [45]. The displacement loop (D-loop) is also a challenging region for both assembly and annotation, as it contains hypervariable sites and a specific three-stranded structure with short linear DNA [46]. Detecting such structures involves repeat detection, regulatory motif identification, and GC content analysis, and automated analysis can help only to some extent. Thus, a wide selection of tools for mitochondrial genome assembly and annotation exists, and selecting the optimal combination for a particular group and data type remains challenging.

1.2. Amphipods as a Model Group to Test and Benchmark Mitochondrial Genome Assembly and Annotation

This study exploits amphipods (Crustacea: Amphipoda) of Lake Baikal as the model group. The order Amphipoda includes about 10,000 formally described species [47] and appears prone to forming species flocks, or groups of closely related species, in ancient lakes [48]. Nuclear genome sizes within amphipods vary more than 100-fold, from 0.52 to 64.62 pg [49,50] (1 pg is equivalent to 0.978 Gb [51]). In mitochondrial genomes of amphipods, the size and gene content are relatively conserved, but gene order variation is quite common (~20%) and can be used as an additional source of phylogenetic signal [52]. Other notable deviations of some amphipod mitochondrial genomes (not unique to them and present in other groups) are connected to tRNA genes and tRNA gene duplication and remolding [53], as well as the presence of aberrant tRNAs lacking one or more conserved structural elements [38]. For these 10,000 species, about 200 mitochondrial genome assemblies for ~155 species [54–95], ranging from 14,113 [57] to 42,256 bp [88], have been published as of 31 August 2026 (Table S1). This sample contains data obtained and assembled with different methods, namely from long-range PCR products, total genomic DNA, and RNA, with Sanger, 454, Illumina, DNBSeq, and Nanopore sequencing. Notably, about half of the assemblies (99) are still incomplete. Moreover, manual assembly at some stage was necessary in many cases [65,68,88,96].

Lake Baikal is the place of origin of at least two species flocks of amphipods [97–99], with a little over 350 officially described species and subspecies and two conflicting taxonomic systems [100]. Known genome sizes in Baikal amphipods vary on the order of 2–17 pg [101], with highly variable repeats [102]. Unsurprisingly, the mitochondrial genomes

are the best-studied part of their genomic systems [100]. At the moment, 18 mitochondrial genome assemblies (13 complete and 5 partial) are publicly available [58,63–65,88,103] (see Table S1). It is important to note that the largest mitochondrial genome in amphipods, which is also one of the largest in animals reported to date, belongs to the Lake Baikal species *Macrohectopus branickii* [88]. So, Lake Baikal amphipods show the same tendencies as amphipods in general, and are even slightly better studied.

Although mitochondrial genomes of amphipods in general and Baikal amphipods in particular are conserved in terms of length and gene content, they seem to present a challenge for assembly. Potential problems that predispose to this challenge may include large genetic diversity with cryptic species [100,104–106], large and variable nuclear genome sizes with abundant repeats [102,107], and nuclear mitochondrial pseudogenes, or NUMTs [108,109]. We therefore decided to benchmark the currently available mitochondrial genome assembly algorithms on the most popular data types (genome skimming and RNA sequencing). Specifically, short reads were used because they are the most prevalent data type in public databases, even though they present the most significant challenges for assembly. Also, we paid special attention to the computational resources required for mitochondrial genome assembly.

The following questions were taken into consideration:

1. Which of the tools (if any) enable reconstruction of the full circular mitochondrial genome without manual finishing, at least for average mitochondrial genomes of ~16 kbp?
2. Should one preferably perform genome skimming or transcriptome sequencing?
3. How much does the depth of sequencing coverage affect assembly quality, and what is the safe minimal depth for assembly?
4. What can be used as a seed when a same-species reference is not available: a *cox1* fragment of the studied species or a mitochondrial genome of a related species?
5. What is the most accurate annotation tool?

In order to answer these questions, we performed assembly for three species with ten assemblers, annotated the resulting assemblies with six annotation pipelines, and recommended a procedure that was suitable for assembling and annotating average mitochondrial genomes of Baikal amphipods on a laptop.

2. Materials and Methods

2.1. Datasets Used for the Analysis

This benchmark was predominantly performed for three species of Baikal amphipods, *Baikalogammarus pullus* (Dybowsky, 1874); *Eulimnogammarus cyaneus* (Dybowsky, 1874); and *Eulimnogammarus verrucosus* (Gerstfeldt, 1858), biological species *W* [110]. These species represent a wide range of nuclear genome sizes, with about 2 pg in *B. pullus*, making it the smallest studied Baikal amphipod genome, about 4 pg in *E. cyaneus*, and about 7 pg in *E. verrucosus W* [101,111].

Genome skimming was performed for DNA extracted from pereopods in the case of larger *Eulimnogammarus* spp. (materials from three individuals pooled in the case of *E. cyaneus* and two in the case of *E. verrucosus W*) and from whole-body material of eight animals in the case of *B. pullus* due to smaller body size. Then, DNA extraction was performed using the DNA Tissue Strong kit (Raissol, Moscow, Russia), following the manufacturer's recommendations from Step 2. Sequencing libraries were prepared with an SG GM kit (Raissol, Moscow, Russia) using the default fragmentation conditions, targeting a fragmentation size of 140–160 bp. The obtained libraries were indexed and pooled for sequencing on a DNBSEQ-G400 platform with paired reads of length 150 bp. DNA extraction, library preparation, and sequencing were performed by the Genomed company

(Moscow, Russia). The raw sequencing reads are available from DDBJ/NCBI/ENA via accession numbers DRR911159, DRR911160, and DRR911166 (BioProject PRJDB40454).

In addition, published RNA sequencing data for the same set of species from different works were used: SRR3467045 for *B. pullus* [99], SRR3467057 for *E. cyaneus* [99] and SRR8205919 for *E. verrucosus W* [112] (37.5 million, 35.4 million, and 30.8 million paired reads, respectively).

To estimate the mitochondrial-to-nuclear sequencing depth ratio, we used publicly available raw read datasets and, where available, mitochondrial genomes from public databases. For *Gammarus nekkensis*, raw reads ERR16722743 were obtained from [113], and the mitogenome was assembled in this study from these reads. For *Hyalella azteca*, raw reads SRR1556042 were obtained from [114], and the mitogenome (GenBank accession MH542433.1/RefSeq NC_039403.1; direct submission, unpublished) was used. For *Parhyale hawaiensis*, raw reads SRR2962758 were obtained from [115], and the mitogenome NC_039402.1 (GenBank accession MH542432.1; direct submission, unpublished) was used. For *Orchestia grillus*, raw reads SRR9871010 were obtained from [116], and the mitogenome was assembled in this study from these reads. For *Trinorchestia longiramus*, raw reads SRR9098169 were obtained from [117], and the published mitogenome NC_039401.1 [78] was used. In addition, the published mitogenome of *Gammarus pisinnus*, NC_044410.1 [86], was used for testing annotation; no raw reads were used for this species. Mitogenomes assembled in this study (for *G. nekkensis* and *O. grillus*) were obtained with MitoFinder, using *G. lacustris* (NC_044469.1, [86]) and *Stygobromus indentatus* (NC_030261.1, [67]), respectively, as references.

Finally, published data for the following Baikal amphipod species was used to test the suggested optimal pipeline: *Eulimnogammarus vittatus* (Dybowsky, 1874) (SRR18803010; [65]), genetic lineage *S* [110]; *Acanthogammarus victorii* (Dybowsky, 1874) (SRR18803012; [65]), *Brachyuropus grewingkii* (Dybowsky, 1874) (SRR18803011; [63]); *Eulimnogammarus verrucosus* (Gerstfeldt, 1858), biological species *S* [111] (DRR911165); *Ommatogammarus albinus* (Dybowsky, 1874) (DRR911170), *Ommatogammarus flavus* (Dybowsky, 1874) (DRR911172); and *Ommatogammarus melanophthalmus* (Dybowsky, 1874) (DRR911171). The mitochondrial genome assemblies for *Ommatogammarus* species were reconstructed in our previous study [118] and are available from NCBI (PZ617559, PZ617560, and PZ617561, respectively), but here we discuss the details of each assembly step.

New mitochondrial genomes assembled in this work are submitted to NCBI GenBank/DDBJ/ENA (accession numbers QB227943 for *B. pullus* and QB180845 for *E. cyaneus*). These genomes and mitochondrial genomes that were reassembled from third-party data are also available from GitHub (https://github.com/drozdovapb/mt_genome_benchmark/tree/main/new_assemblies; accessed on 22 November 2026).

2.2. Tools Used for Mitochondrial Genome Assembly

The main analysis performed in this work was a benchmark of automated mitochondrial genome assembly tools. The following criteria needed to be met by the tools to be included in the benchmark: (1) the tool is intended for mitochondrial genome assembly; (2) suitable for short reads; (3) available for download at the moment; and (4) can be installed. This gave us the list of ten tools, namely ARC [32], GetOrganelle [28], MEANGS [27], MITGARD [25], MITObim [24], MitoFinder [16], MitoZ [31], mtGrasp [26], Norgal [29], and NOVOPlasty [30]. They employ a wide range of strategies and assemblies (Table 1). In addition, we tried but could not run IOGA [119], mitoRNA [20], MitoFlex [33], and MitoMaker [34], which are not maintained. We did not include Trimitomics [120] because it is a textual description of an example assembly procedure rather than an automatic

pipeline. Finally, OrgaMiner [36] and skim2mito [37] were also not included in the benchmark because they reuse GetOrganelle, which was already included.

Table 1. Assembly tools used in this work. After each repository address, the year of the most recent commit is provided as a proxy for whether the tool is maintained (all accessed on 31 July 2026). Availability in Galaxy [121] last checked as of 13 July 2026.

Tool/Version Used	Approach	Assembler	Reference/Available From (Last Updated)
ARC 1.1.4-beta	Mapping, splitting, assembly, and finishing	Newbler or SPAdes [122]	[32] https://github.com/ibest/ARC.git (2016)
GetOrganelle 1.7.4.1	Baiting and iterative mapping; de novo assembly	SPAdes [122]	[28] https://github.com/Kinggerm/GetOrganelle (2025)
MEANGS 1.0	nhmmer [123] to extract reads; string overlap with SSAKE; seeds for de novo assembly	SSAKE [124]	[27] http://github.com/YanCCscu/meangs (2026)
MITGARD 1.0	Alignment to reference; de novo assembly Specific to RNA-seq	Trinity [125], rnaSPAdes [126], MitoZ [31]	[25] https://github.com/pedronachtigall/MITGARD (2024)
MITObim 1.9.1	Baiting and iterative mapping	MIRA [127]	[24] https://github.com/chrishah/MITObim (2018) https://usegalaxy.org.au; https://usegalaxy.eu
MitoFinder 1.4.2	De novo assembly; finding mitochondrial contigs	MEGAHIT, MetaSPAdes, or IDBA [128–130]	[16] https://github.com/RemiAllio/MitoFinder_container/ (2024) https://github.com/RemiAllio/MitoFinder (2025)
MitoZ 3.6	De novo assembly	SOAPdenovo-Trans [131]	[31] https://github.com/linzhi2013/MitoZ (2024)
mtGrasp 1.18	De novo	ABySS [132]	[26] https://github.com/bcgsc/mtGrasp (2026)
Norgal 1.0.0	De novo assembly of binned reads	MEGAHIT [127] and IDBA [130]	[29] https://bitbucket.org/kosaidtu/norgal (2018)
NOVOPlasty 4.3.5	Seed-and-extend	Own algorithm	[30] https://github.com/ndierckx/NOVOPlasty (2024) https://usegalaxy.fr; https://usegalaxy.org.au; https://usegalaxy.eu

Each tool was installed according to the manual and tested on an identical set of reads. Reads were preprocessed by adapter and quality trimming with trimmomatic v0.39 [133] (trimming parameters: ILLUMINACLIP:Seq_adapters.fasta:2:7:1) for the main benchmark and with bbduk.sh from BBTools v38.95 [134] (trimming parameters: ktrim=r k=23 mink=11 hdist=1) for the recommended pipeline. Random subsampling of raw reads was then performed with the sample command from seqtk v1.3-r106 [135] in two scenarios. First, to eliminate the confounding effect of unequal sequencing depth between species in the main assembly benchmark, the DNA datasets of *E. cyaneus* and *B. pullus* were downsampled to approximately 2× nuclear genome sequencing depth, matching that of *E. verrucosus W*. Second, to evaluate the effect of sequencing depth on assembly quality,

additional random subsets were derived from the DNA datasets of *H. azteca* and *E. cyaneus*, spanning 0.5×–0.005× and 3.8×–0.04× nuclear genome sequencing depths, respectively. In both cases, subsampling was applied before assembly, and all assemblers received the same subsampled reads to ensure a fair comparison. Then, all assemblers were run using custom wrapper scripts.

The reference sequences for the tools that require them were obtained as follows. The reference sequence for *E. cyaneus* was the published NC_033360/KX341964 assembly, which was obtained with short-read sequencing of long-range PCR amplicons [65]. The reference sequence for *E. verrucosus W* was the published NC_023104.1/KF690638 assembly, which was assembled with a crystallization strategy of iterative mapping to a set of distantly related references at a low mapping quality threshold and subsequent reference mutation [103]. The reference sequence for *B. pullus*, which was assembled with Trinity, was retrieved from NCBI GenBank as BK059231 [58]. The *E. verrucosus W* assembly served as a related reference for *E. cyaneus* and *B. pullus*, and the *E. cyaneus* assembly was used as a related reference for *E. verrucosus W.* The Folmer *cox1* fragments for *B. pullus* (FJ756303), *E. cyaneus* (MK887733) [110], and *E. verrucosus W* (MK887402) [110] were retrieved from GenBank. The Folmer fragments in fasta can be converted to the GenBank format required for MitoFinder using the script gb_creator_from_fa.py available in our repository (https://github.com/drozdovapb/mt_genome_benchmark/blob/main/1_assembly/Developed_tools/gb_creator_from_fa.py; accessed on 21 September 2026).

The quality of the resulting assemblies was assessed based on their similarity to the reference assembly and gene content. Similarity to the reference, i.e., the quality of the reference reconstruction, was estimated using the alignment-based score suggested by [136], which consists of four aspects, each contributing up to 25% of the total score, namely the number of contigs ($n_{contigs}$, ideally one), coverage of the reference by the assembly (cov_{qry}), coverage of the assembly by the reference (cov_{ref}) and repeat resolution as follows:

$$\text{score} = ¼ \times (cov_{ref} + cov_{qry} + \min\{(cov_{ref}/cov_{qry}), (cov_{qry}/cov_{ref})\} + 1/n_{contigs}) \quad (1)$$

The calculation of the score employed minimap2 v2.24-r1122 [137] and bedtools v2.30.0 [138]. Completeness was counted as the number of protein- and rRNA-coding genes predicted by MitoFinder in the annotation mode (-a) using the reference genome of the corresponding species. The expected number of genes is 15 (13 PCGs and 2 rRNA genes). Also, the resources required for running each tool and data combination were recorded using a custom bash script (https://github.com/drozdovapb/mt_genome_benchmark/blob/main/1_assembly/Developed_tools/monitor_PPID2407_2.sh; accessed on 29 October 2026).

Estimation of coverage depth was performed according to the following procedure**.** Mitochondrial read depth: Paired reads were mapped to reference mitochondrial genomes with BWA-MEM v0.7.12-r1039 [139] (bwa mem -t 4 <reference.fasta> <reads_R1.fq.gz> <reads_R2.fq.gz>); secondary and low-quality alignments (MAPQ > 20) were filtered out with SAMtools v1.9 [140] (samtools view -bS -F 0x100 -q 20), and mean mtDNA read depth was calculated using samtools depth across all reference positions, including uncovered ones. Nuclear read depth: Because complete nuclear genomes are not available for these amphipod species, nuclear coverage depth was estimated analytically as

$$CD_{nuc} = (N_{reads} \times L_{read} \times 2)/GS_{nuc}, \quad (2)$$

where N_{reads} is the number of reads in one file, L_{read} is the read length, the factor of 2 accounts for paired-end data, and GS_{nuc} is the haploid nuclear genome size obtained from the literature [101,111,113–117]; original datasets were subsampled with the seqtk sample command from seqtk v1.3-r106 [135], using the same random seed for forward and reverse

read files. The primary metric was CD_{mt}/CD_{nuc}, or coverage depth ratio index. Circular genome visualization: Circular maps of mitochondrial genomes were generated with a custom Python v3.10.12 script, circular_mt_plot.py (https://github.com/drozdovapb/mt_genome_benchmark/blob/main/article-figures/scripts/python/circular_mt_plot.py; accessed on 29 October 2026). Gene annotations were parsed from GenBank files using Biopython v1.83 [141], with tRNA features excluded from the visualization. Per-position coverage depth was computed independently from the indexed BAM files with the pysam library v0.22 [142], which wraps SAMtools [140] and htslib [143]. The BAM files used as input were produced by the companion script run_mt_coverage.py (https://github.com/drozdovapb/mt_genome_benchmark/blob/main/article-figures/scripts/python/run_mt_coverage.py; accessed on 29 October 2026). Numerical operations were performed with NumPy v1.26 [144], and figures were rendered with Matplotlib v3.8 [145]. Final circular genome maps are available in the Supplementary Materials.

Identification of putative NUMTs was performed in the following way. To test whether the short duplicate *nad4* + *nad5* contig in *O. flavus* could represent a nuclear mitochondrial insertion, we mapped the raw reads to a combined reference containing both the main mitochondrial contig and the short duplicate contig using BWA-MEM in the same way as for the sequencing depth estimation (see above). Sequence divergence between the two contigs was assessed using BLASTN [146] with default parameters. The mean sequencing depth and mapping quality (MAPQ) for each contig were calculated using SAMtools (v1.9), as described above. To evaluate alignment competition caused by cross-mapping, we calculated the proportion of soft-clipped reads (reads with 'S' in their CIGAR string) using samtools view -F 0x100 -q 20. All commands and parameters are provided in the GitHub repository https://github.com/drozdovapb/mt_genome_benchmark/tree/main/article-figures/scripts/bash (accessed on 22 November 2026).

2.3. Tools Used for Annotation

Several annotation pipelines were tested. Here, we applied AGORA [147], GeSeq [148], MITOS2 [149], and MFannot [150], as well as the annotation modules of MitoZ [31] and MitoFinder [16]. The tested pipelines employ a variety of approaches (Table 2) but combine very similar sets of standalone tools (Table S3). For protein-coding gene annotation, BLAST- [146], BLAT- [151] and hidden Markov model (HMM)-based tools [152,153] are most commonly used. rRNA gene annotation uses either BLAST-based approaches or a covariance model (CM) approach, as well as most of the tools for tRNA annotation. CMs include sequence and structural information about one specific tRNA type (or other molecule with stable secondary structure). In some cases, a curated initial dataset is needed, and clade-specific CM sets can be prepared. MiTFi [154] and tRNAscan [155], the most widely used in annotation pipelines, are based on the Infernal Suite implementing CMs [156], which can also be used as a standalone annotation tool. Approaches including heuristic tRNA pattern search, ARWEN [157] and ARAGORN [158] are based on stable tRNA structure detection.

Table 2. Annotation pipelines used in this work. After each repository address, the year of the most recent commit is provided as a proxy for whether the tool is maintained (all accessed on 31 July 2026). Availability in Galaxy [121] was last checked as of 15 July 2026.

Tool/Version Used	Algorithm			Reference/ Available from (Last Updated)
	PCGs	rRNAs	tRNAs	
AGORA (accessed 21 September 2026)	tblastn	BLASTN	BLASTN	[147]

https://doi.org/10.3390/d18120741

				https://datalab.dongguk.edu/gene_project/AGORA/
GeSeq (accessed 21 September 2026)	BLAT + HMM	BLAT + HMM	BLAT + tRNAscan-SE (COVE mode) + ARAGORN	[148] https://chlorobox.mpimp-golm.mpg.de/geseq.html
MFannot (accessed 21 September 2026)	BLAST, exonerate, RNAweasel, HMMER (intron-aware)	RNAweasel (uses Infernal or ERPIN)	RNAweasel (uses Infernal or ERPIN)	[150] https://megasun.bch.umontreal.ca/apps/mfannot/
MITOS2 2.1.9	HMM or BLASTX (RefSeq) + start/stop refinement	CM, based on Infernal (MiTFi and tRNAscan-SE)	CM, based on Infernal (MiTFi and tRNAscan-SE)	[149] https://gitlab.com/Bernt/MITOS/-/tree/mitos2 (2025) https://usegalaxy.org https://usegalaxy.eu https://usegalaxy.org.au https://usegalaxy.fr
MitoFinder 1.4.2	In-house Python script (BLAST-based), by reference	In-house Python script (BLAST-based), by reference	MiTFi/ARWEN/tRNAscan-SE	[16] https://github.com/RemiAllio/MitoFinder_container/ (2024) https://github.com/RemiAllio/MitoFinder (2025) (accessed on 31 July 2026)
MitoZ 3.6	In-house Perl script	MiTFi	Infernal	[31] https://github.com/linzhi2013/MitoZ (2024)

Assembly of *E. cyaneus* obtained with MITObim was used as a test case to compare different annotation approaches. In the AGORA web application, two references were used: *E. cyaneus* (accession: NC_033360.1) and *G. pisinnus* (accession: NC_044410.1). The invertebrate mitochondrial genetic code (5) was selected. The GeSeq web application allows annotation using not only a reference species, but also a broader taxonomic group: we selected the genus *Gammarus*. The tool was configured as follows: the mitochondrial genome was specified as circular, the default BLAT parameters were retained, and all three available tRNA prediction tools (ARAGORN, ARWEN, and tRNAscan-SE) were selected using parameters appropriate for mitochondrial tRNA annotation. For MFannot, the invertebrate mitochondrial genetic code (translation table 5) was selected.

The MITOS2 pipeline was used for de novo mitochondrial genome annotation as a command-line tool, but it is also available in Galaxy [121]. The MITOS2 developers provide curated model sets for tRNAs, rRNAs, proteins, and both O_H and O_L regions based on different versions of the RefSeq database. We evaluated two reference datasets, RefSeq 63 and RefSeq 89. For additional tRNA annotation with Infernal, CMs for tRNAs from RefSeq 89, provided by the developers of MITOS2, were used as a general (default) set, while Amphipoda-specific CMs, published by Romanova et al. [53], were used as a clade-specific set.

Both MitoZ and MitoFinder can provide mitochondrial genome annotation in addition to the assembly. Here, we encountered two technical difficulties when annotating the assemblies produced with some tools using MitoFinder. The first issue concerned detecting circularity in mitochondrial genomes. The main scripts of the MitoFinder annotation pipeline run under Python 3 (the version used in our base environment), whereas the circularity check script (circularizationCheck.py) requires Python 2. Because of this version conflict, MitoFinder reported "Circularization: Not found" in the statistics file (.infos) after annotation, even though circularization was successfully detected when circulariza-

tionCheck.py was run separately with Python 2. To avoid this problem, we strongly recommend installing MitoFinder in an isolated conda environment with the correct Python version. The second issue was the sensitivity of MitoFinder to contig name length. We discovered this by jointly analyzing the assembly score values, contig lengths, and the number of annotated genes. We were particularly interested in assemblies that had non-zero scores but contained no annotated genes: the assembly aligned to the reference, yet the genes were not annotated. Examination of the annotation logs for these assemblies revealed that MitoFinder simply reported that no mitochondrial contigs were found. Drawing on our experience with tools that are sensitive to sequence-identifier length, we hypothesized that the cause lay specifically in long contig names, and we successfully annotated the majority of the problematic assemblies after shortening the names. We therefore recommend that, before using MitoFinder as an annotator, contig names in the input fasta file should be adjusted appropriately.

Finally, some pipelines were not tested. DOGMA [159] and MOSAS [160] were not available, and we failed to run the annotation modules of mitoconstrictor [35] and MitoFlex [33]. These pipelines use some combinations of the same tools as the ones we were able to test (Table S4). In contrast, DeGeCI (de Bruijn graph gene cluster identification) uses a different approach, but we could not run it either: we failed to install the standalone version [161], and the web version [162] was also unavailable (https://degeci.informatik.uni-leipzig.de/; accessed on 21 September 2026).

The RefSeq annotation for *E. cyaneus* (accession: NC_033360.1) was used as the ground truth for this work. To independently assess the accuracy of the annotated reference proteins, we additionally validated the end coordinates of corresponding genes. Prediction of putative RNA cleavage ratios at gene boundaries was performed using *E. cyaneus* RNA-seq reads from NCBI BioProjects PRJNA319818 [99] and PRJNA871894 (seven samples in total), following the methodology published in [163] and the code provided by the authors. Quality-trimmed reads were polyA-trimmed with PRINSEQ-lite v0.20.4 [164] and aligned to the *E. cyaneus* mitochondrial genome assembly using STAR v2.7.10b [165]. Then, cleavage ratios and peak sites were identified using the scripts provided in the repository (https://github.com/gcarbajosa/Mitochondrial_RNA_Cleavage; accessed on 31 July 2026). The read depth was calculated from the BAM files using bedtools v2.30.0 [138] and then log-normalized; then, the median depth was calculated across the samples.

The reference annotation was moved to the new assembly using BLASTN (2.13.0). Differences from the reference annotation were analyzed by comparing coordinates. If the start and end coordinates matched completely or differed by less than 10 bp, the gene was considered a true positive. If the gene was found but coordinates differed by 10 bp or more, or the gene was found in multiple locations, it was considered a false positive. If the gene was not found, it was considered a true negative. For cases where a gene was found and its coordinates intersected with those in the reference annotation, the deviation in coordinates was calculated as the sum of the absolute deviations of the start and end coordinates. This value was calculated for each annotation attempt across all gene annotations intersecting the reference ones, and the mean deviation was used to characterize the annotation pipelines. Visualization of the gene annotations in the coordinates of the *E. cyaneus* mitochondrial genome assembly was performed in R using the ggplot2 v4.0.3 [166], gggenes v0.7.0 [167], and readr v2.2.0 [168] packages and finalized with Inkscape v1.1.2 (https://inkscape.org; accessed on 22 November 2026). The presence/absence gene matrix design for genes was inspired by [169].

2.4. Suggested Procedure for Assembly and Annotation of a New Mitochondrial Genome

We found that the following combination of tools provided optimal results while minimizing the number of software tools needing installation. This procedure is oriented

toward short-read lcWGS data, needs a suitable annotated reference mitochondrial genome, and consists of the following steps. First, the reads were quality- and adapter-trimmed with bbduk.sh from the BBTools suite [134] (version 38.95 was used) with parameters and adapter files suggested by the BBTools manual. Second, the reads were optionally selected based on k-mers of the reference sequence using bbduk (k = 17 rcomp = t qhdist = 0) from the same BBTools suite, which reduces computational load. Then, the genome was assembled and annotated with MitoFinder [16] (version 1.4.2 was used). In cases of incomplete (not circular) assemblies, they were further elongated with MITObim [24] (version 1.9.1 tested) using the MitoFinder assembly (or the largest contig from it) as a reference. Then, the MITObim assemblies were again annotated with MitoFinder, and whether the annotation lacked some expected genes, and also with MITOS2.

The current version of the procedure, along with reproducible code for all analyses used in this manuscript, is available on GitHub at https://github.com/drozdovapb/mt_genome_benchmark (accessed on 22 November 2026).

3. Results

3.1. Assembly Benchmark

We started by comparing the performance of ten assembly pipelines in ideal circumstances, i.e., when the complete mitogenome of the same species was supplied as a reference whenever applicable. For that, we used three species with different levels of complexity in their nuclear genomes: *B. pullus* with a genome size of about 2 pg (~2 Gbp) [101]; *E. cyaneus* with a genome size of about 4 pg (~4 Gbp) [101]; and *E. verrucosus W* with a genome size of about 7 pg (~7 Gbp) [111]. All datasets were resampled to an approximate nuclear genome read depth of 2×, corresponding to 14.4, 24.4, and 41.8 million reads, respectively. We found that in the case of *B. pullus* the majority of the assemblers used, six out of ten, successfully recovered 15 genes (13 PCGs and 2 rRNA genes), with scores close to 100%; three more assemblers recovered 15 genes, but not complete genomes; and the remaining assembler recovered 13 genes (Figure 1A). For the two species with larger genomes, only MITObim yielded an assembly with a score of 100%, and two other assemblers (MITGARD and ARC) managed to recover all 15 genes for *E. verrucosus W* (Figure 1B,C).

Also, we compared lcWGS to RNA-seq (31–37 million reads) of the same species. In general, RNA-seq performed worse. However, MitoZ achieved scores close to 100% and recovered 15 genes for two species, and MITGARD recovered 15 genes for all three species (Figure 1G–I). In addition, Norgal, MITObim and MitoFinder recovered all 15 genes in some cases, but the results were not consistent across species.

The next question is the depth of coverage. It is an open question what read depth is needed to successfully extract a full, circular mitochondrial genome from short-read data, while it is also an important thing to consider when planning a genome project. However, it is logical to suggest that the defining value would be the actual read depth of the mitochondrial genome rather than the coverage depth of the nuclear genome. As there were no estimates for mitochondrial to nuclear coverage depth ratio for amphipods, we estimated it for eight species and found that it varies by around 9-fold in this sample (Figure 2A). Then, we chose two contrasting specimens, those of *H. azteca* with the largest ratio in our sample and *E. cyaneus* with one of the lowest, and compared all ten assemblers on sets of subsampled reads. These sets ranged from 0.5× to 0.005× nuclear genome read depth for *H. azteca* and 3.8× to 0.04× for *E. cyaneus* (Figure 2B). The results of this analysis showed progressive worsening of the assembly with MITObim, showing the best results at a low read depth.

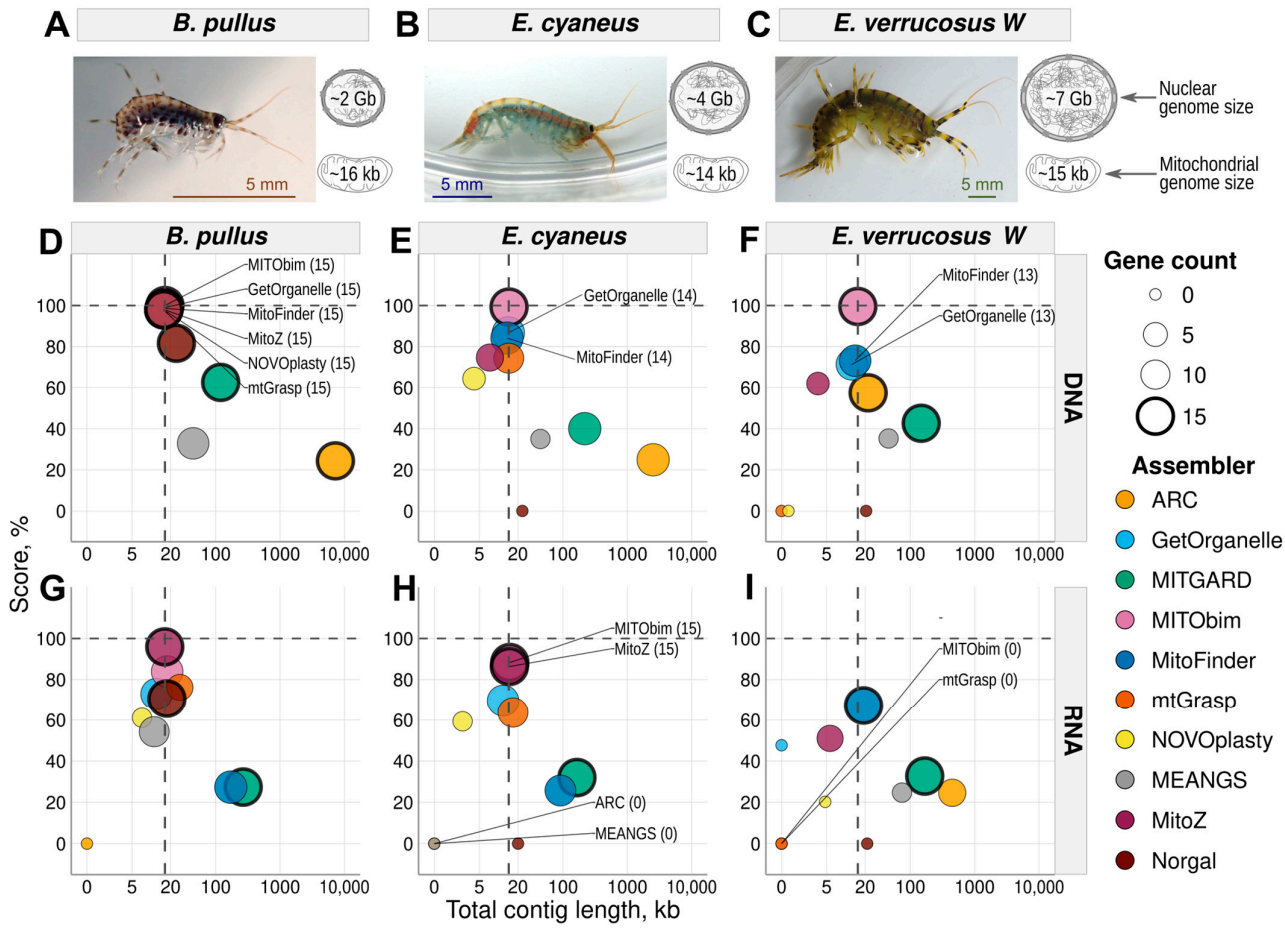


Figure 1. Quality of mitochondrial genomes assembled for three species with different nuclear genome sizes (in columns) based on DNA or RNA sequencing. (**A**–**C**) Typical representatives of the studied species and their nuclear and mitochondrial genome sizes. (**D**–**F**) Benchmark of all assemblers on DNA sequencing data and (**G**–**I**) benchmark of all assemblers on RNA sequencing data. All overlapping points are additionally labeled. The gray dashed lines correspond to the target values of mitochondrial genome size and assembly score: the horizontal line corresponds to a 100% score, while the vertical line is placed at the reference genome size values for each species, i.e., 16,228 bp for *B. pullus*, 14,370 bp for *E. cyaneus*, and 15,315 bp for *E. verrucosus W*. All assemblers were run with default parameter sets (ARC used SPAdes, MitoFinder was run with MEGAHIT, and MITGARD used Trinity and rnaSPAdes), and the same-species mitogenome reference was provided as the bait, reference, or seed sequence for all pipelines that required those. See Table S5 for the characteristics of each assembly.

Previous iterations showed that the best scores that were consistent between species were yielded by MITObim, which is highly dependent on the seed. However, such an ideal seed as the mitogenome of the same species is rarely available, and most challenges arise in the projects exploring novel and non-model organisms without a same-species reference. However, at the same time, the availability of a short marker sequence of the 5′ half of the *cox1* (Folmer region) is much more realistic. Thus, we checked whether each of the seed-based tools still works using related mitogenomes or a Folmer region from the same species. We see that while, for some tools, the performance is diminished with decreasing length of the reference, many tools are not visibly affected (Figure 3). MitoZ,

MEANGS, and Norgal were not tested, as they do not depend on the provided seed/reference, and the results were provided in Figure 1. The success of using a *cox1* fragment vs. a related mitogenome varied and depended on the closeness of the reference. While in the case of *B. pullus*, the reference was more distant and the *cox1* fragment worked better, for the other two species, *E. cyaneus* and *E. verrucosus W*, for which congeneric species were used as references, the whole-genome reference of the other species should be preferred. Here, it is very important to note that MITObim produced artifacts when provided with a related mitogenome. In the case of *B. pullus* mitogenome assembly using *E. verrucosus W* mitogenome as a reference, the result was identical to the reference.

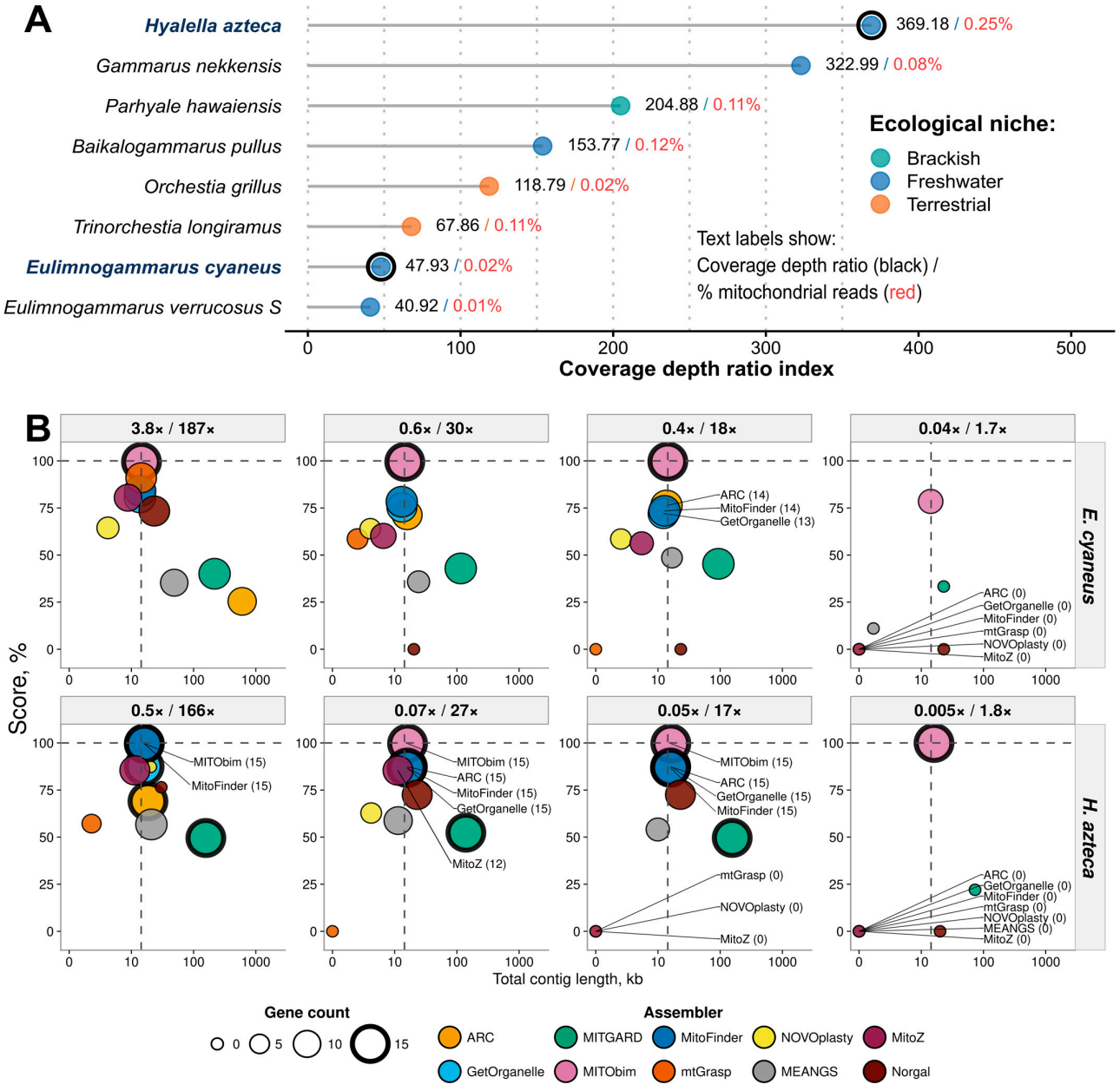


Figure 2. Effect of the depth of sequencing coverage on the quality of mitochondrial genome assembly. (**A**) Nuclear/mitochondrial coverage depth ratios and percent of mitochondrial reads in nine sequencing libraries of different amphipod species. (**B**) Quality of mitochondrial genome assemblies obtained from reads of *E. cyaneus* (top row) and *H. azteca* (bottom row), providing different read depths. Horizontal panel titles indicate nuclear and mitochondrial genome read depth for each set

https://doi.org/10.3390/d18120741

of reads. Total contig length is the sum of lengths for all contigs or scaffolds (whichever was yielded by the particular assembler). Please note that the horizontal axis in (**B**) is logarithmic. The gray dashed lines correspond to the target values for mitochondrial genome size and assembly score: the horizontal line corresponds to a 100% score, while the vertical line is placed at the reference genome size values for each species, i.e., 14,370 bp for *E. cyaneus* and 15,991 bp for *H. azteca*.

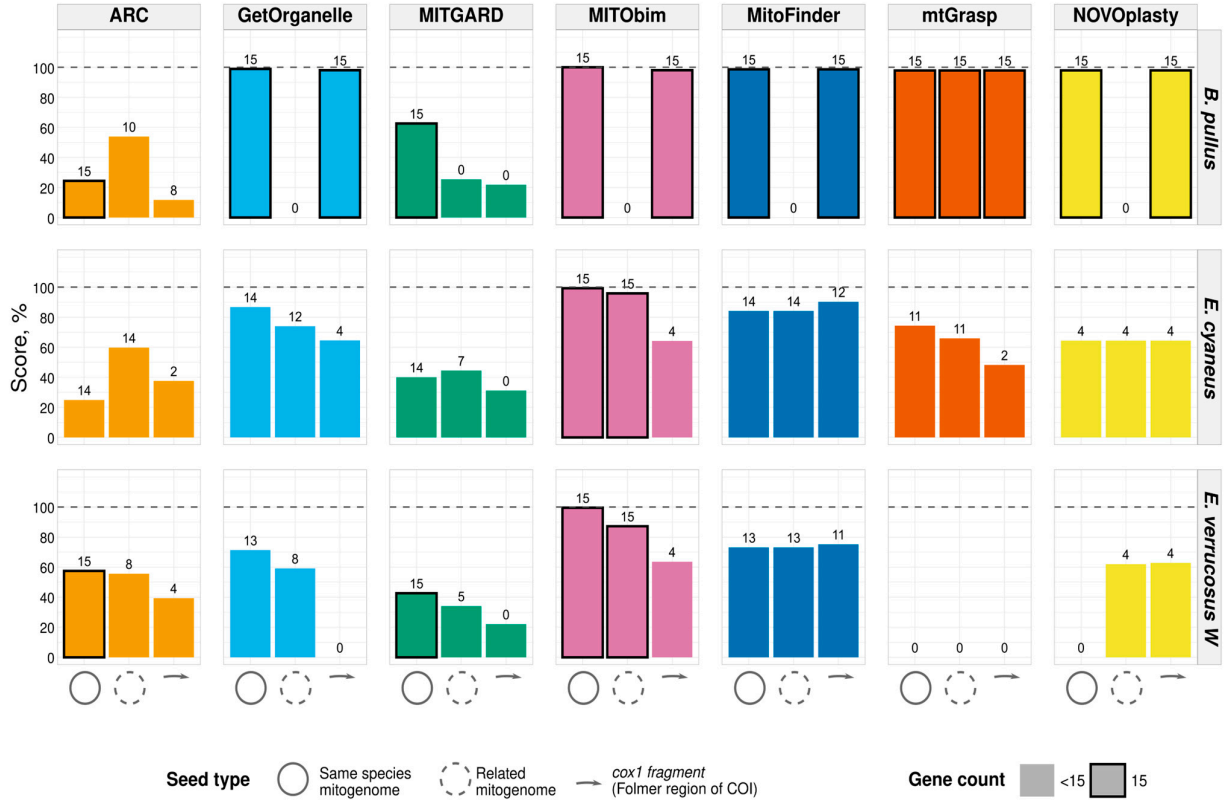


Figure 3. The effect of reference/seed choice on mitochondrial genome assembly quality. The numbers above the bars correspond to the number of genes (out of 15 searched) found in the corresponding assembly with MitoFinder. See Figure S1 for the information on total contig length.

Last but not least, an important consideration for a project is the computational resources (mostly time and memory needed to run a particular assembly). Here, we compared the resources that were needed for a set of 96 assemblies (see Table S5). This set consisted of 72 assemblies from DNA data resampled to ~2× nuclear genome read depth (3 species × (7 seed-based assemblers × 3 types of seeds + 3 de novo assemblers)) and 24 assemblies for *E. cyaneus* RNA data (7 seed-based assemblers × 3 seeds + 3 de novo assemblers). Peak memory consumption in terms of resident set size varied from 116 MB to 60 GB but mostly did not exceed 10 GB, with notable exceptions being NOVOplasty (Figure 4). Analysis time also ranged from under 1 min to over 32 h but mostly (94 out of 96 cases) stayed below 12 h, which could be a reasonable overnight run. Importantly, the longest runs, which also required large memory allocation, were those on RNA sequencing data by MITOtim that did not produce meaningful results (see Figure 1G–I). At the same time, DNA-based MITObim assemblies, which were one of the best in terms of quality (see Figure 1D–F), required very little memory allocation and had reasonable run times (Figure 4).

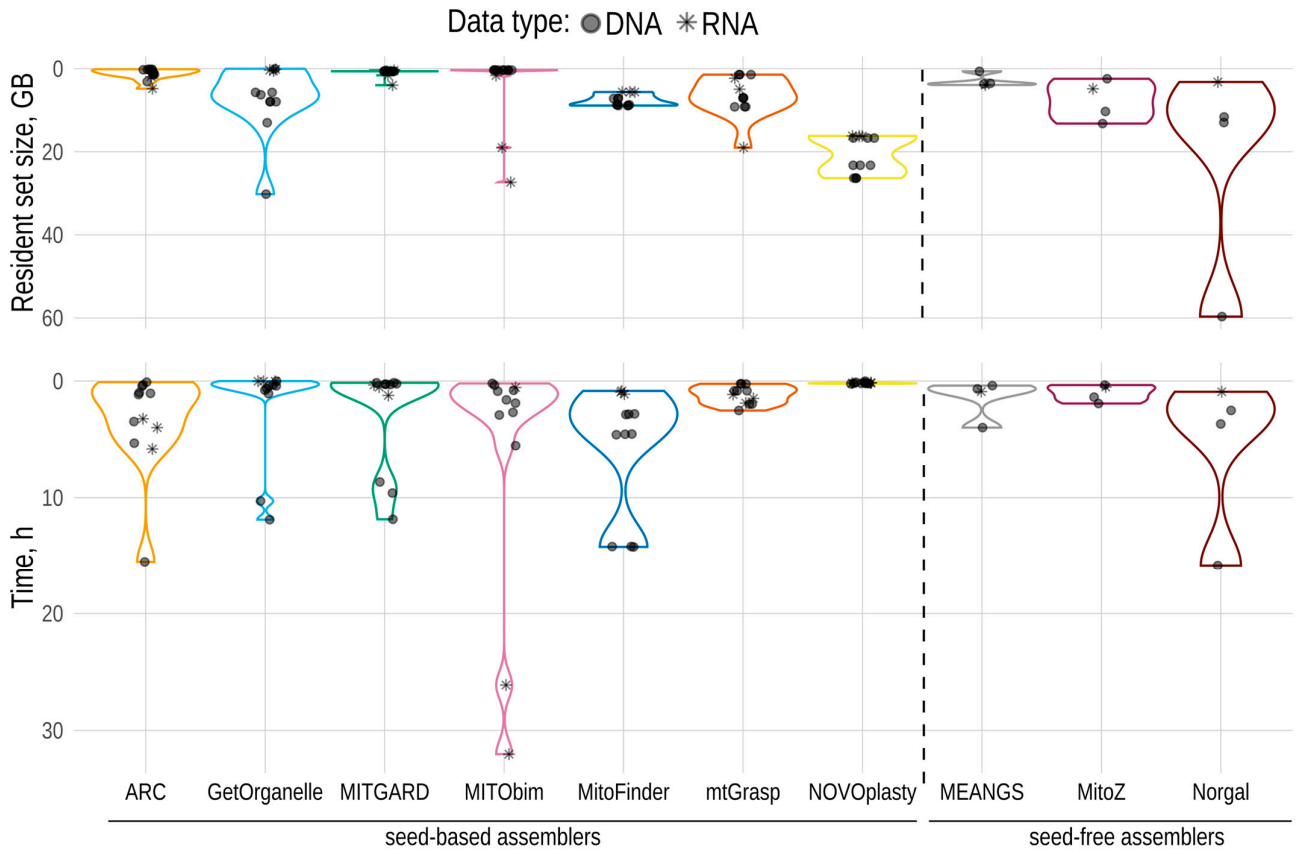


Figure 4. Resource usage by different assemblers. Each individual point represents a separate assembly attempt (circles and asterisks stand for DNA and RNA sequencing reads, respectively), and the violin plots show the overall distribution shape. Please note that the vertical axis is reversed because values closer to zero are better (faster assembly or lower memory requirements).

3.2. Annotation Benchmark

At the next step, we compared different annotation pipelines using the new *E. cyaneus* assembly, as for this species we had a high-quality RefSeq assembly obtained with long-range PCR and annotated manually [65], and our assembly was ~98% identical to it. Analysis of the prediction of putative RNA cleavage ratios using RNA-seq data supported annotation coordinates for *cox1*, *cox2*, *nad2*, and *nad4* genes that were annotated to have incomplete stop codons (Figure S2). Then, we used the transferred RefSeq annotation as the gold standard and compared annotation pipelines to it (Figure 5).

First, we compared general CMs, provided by MITOS2 developers, and Amphipoda-specific CMs, provided by Romanova et al. [53], using Infernal (Figure 5B). As expected, the Amphipoda-specific CMs recovered an identical set of 22 tRNA genes in our assembly, as the annotation transferred from the RefSeq assembly. General CMs also produced 22 tRNA genes, but *trnV* was missing, *trnA* was duplicated, and *trnG* and *trnY* were recognized as *trnW* and *trnC*, respectively (Figure 5B,D; Table S6). Most of the detected mitochondrial tRNAs were expected to have a standard cloverleaf structure, with two exceptions. First, one copy of the *trnS* gene lacks the D-arm in metazoans [38]. The *trnV* gene, on the other hand, shows a tendency to lose the D-arm specifically in the Amphipoda clade [65]. Indeed, most tRNAs were predicted to have a standard cloverleaf structure, with the exceptions being *trnS1* (D-armless in both predictions) and *trnA* (T-armless in the prediction with general CMs). Interestingly, the *trnV* gene was missing from the general CMs prediction, but with Amphipoda-specific CMs it was predicted to have a cloverleaf

structure. However, a closer inspection showed that the D-arm loop consisted of only one base pair, so the tRNA can be interpreted as D-armless. It could be linked to the Amphipoda-specific *trnV* model, which includes a variety of tRNA shapes across species in the clade and can report loops even from a small number of bases.

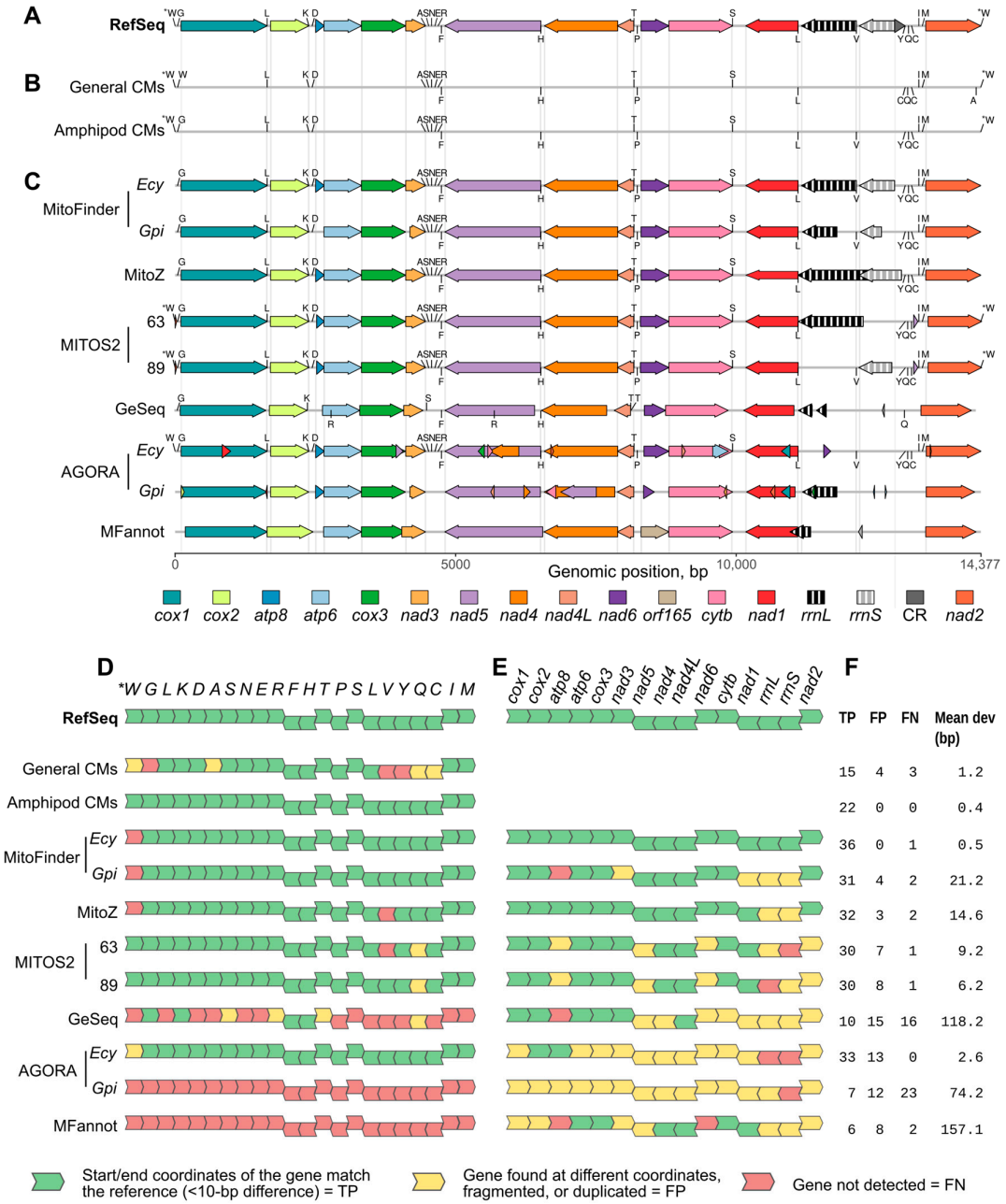


Figure 5. Comparison of annotation tools using *E. cyaneus* mitochondrial genome assembly as an example. (**A**) Annotation from the RefSeq assembly of the same species in the coordinates of an assembly obtained in this work. (**B**) Comparison of tRNA annotations obtained with general and Amphipoda-specific CMs. (**C**) Comparison of annotation pipelines. Reference species or database is indicated in parentheses wherever applicable. tRNA genes are shown with lines at the left coordinates, with one-letter designations for tRNA families next to them. tRNA genes marked with an asterisk represent one divided gene that crosses the zero coordinate of the assembly. Protein-coding and rRNA genes are represented by thick colored and striped lines with arrows, respectively, and

the legend is below. The borders of the PCGs and rRNA genes in the RefSeq assembly-based annotation are drawn with light gray lines for easier visual correspondence. (**D**,**E**) Detection summary for tRNA genes (**D**), rRNA genes and PCGs (**E**), and descriptive statistics (**F**) for each annotation attempt. TP, true positive; FP, false positive (significantly shifted coordinates and duplicated genes); FN, false negative (gene not found or assigned to a different family); Mean dev (bp), mean value of the sum of absolute deviations of start and end coordinates for each found gene. Coordinates for each annotation and all intermediate calculations are provided in Table S6.

Two of the assembly pipelines, MitoFinder and MitoZ, can also be used for annotation. MitoFinder uses the user-provided sequences of PCGs and rRNA genes in a reference GenBank file and predicts tRNA genes de novo. Here, two references were used: the published assembly for the same species as a nearly identical reference and the published assembly for *G. pisinnus* as a distantly related one. The annotation produced by MitoFinder using the *E. cyaneus* RefSeq assembly as a reference was very similar to the RefSeq annotation in terms of PCGs and rRNA genes; in the case of tRNA, 21 genes were predicted, but *trnW* was missing. The annotation using *G. pisinnus* as a reference was clearly worse, as it missed the *atp8* gene and had shorter rRNA gene annotations; the tRNA annotations were identical to those in the previous result because they are produced de novo. The annotation produced by MitoZ recovered very similar PCG positions to those in RefSeq and MitoFinder annotations; however, rRNA gene positions differed from the RefSeq annotation and were also overlapping, which is implausible, and 20 tRNA genes were predicted (*trnW* and *trnV* were missing).

MITOS2 is a specialized pipeline for automated annotation of metazoan mitochondrial genomes that uses its own curated database based on RefSeq and does not require particular reference annotations to be provided. It was run using two versions of the database based on RefSeq63 and 89, respectively. Both returned identical sets of PCGs, which included all 13 expected genes (coordinates for 12 were identical) and an additional false-positive fragment of *nad5*. The set of tRNA genes was identical to the *E. cyaneus* RefSeq annotation only when MITOS2 used the database based on the RefSeq89 version, while the annotation based on RefSeq63 missed the D-armless *trnV* gene, and surprisingly, each of these attempts returned only one of the two rRNA genes, either *rrnL* or *rrnS*. MITOS2 is also the only tool in our benchmark that can annotate the control region. However, the search for CR elements failed, as no hits for the O_L region were found with the CM approach, and no hits were found for O_H with a similarity search.

Finally, three general-purpose annotators, GeSeq, AGORA, and MFannot, were used. GeSeq was developed for plants but can be used for animals if an appropriate reference is provided [148], and the genus *Gammarus* was indicated as the target group. AGORA can be used for any mitochondrial or plastid eukaryote genome; an appropriate specific reference should also be provided [147], and it was tested using the RefSeq *E. cyaneus* annotation as a same-species reference and the *G. pisinnus* annotation as a distantly related reference. MFannot is optimized for eukaryotes other than bilaterian animals [150], so our group was outside its taxonomic scope, but the analysis was still run for comparison. The results show that none of these pipelines annotated exactly 13 protein-coding genes, and annotation of rRNA and tRNA genes was either absent or significantly wrong (Figure 5C).

3.3. Suggested Reference-Based Procedure for Assembly and Annotation

Overall, we tried 454 combinations of data set, reference and assembler for the three studied species (Table S5). Of these, only two combinations produced a score of 100% and 15 annotated genes, and both were MITObim assemblies using a same-species reference. In the case of RNA sequencing, only MitoZ gave reasonable results for different species (Figure 1), so it was not compared with other options. In order to more directly compare

the assemblers for DNA data, we used the same 72 assemblies as described for the resource-usage analysis. From this set, MITObim again returned the highest median score, followed by MitoFinder and MitoZ (Figure 6A). Results are more positive if the intended output is to recover mitochondrial genes, not necessarily a complete assembly: 23 out of 72 assemblies had the 15 genes. Leading assemblers again included MITObim and MitoFinder with medians of 15 and 14 annotated genes, respectively (Figure 6B).

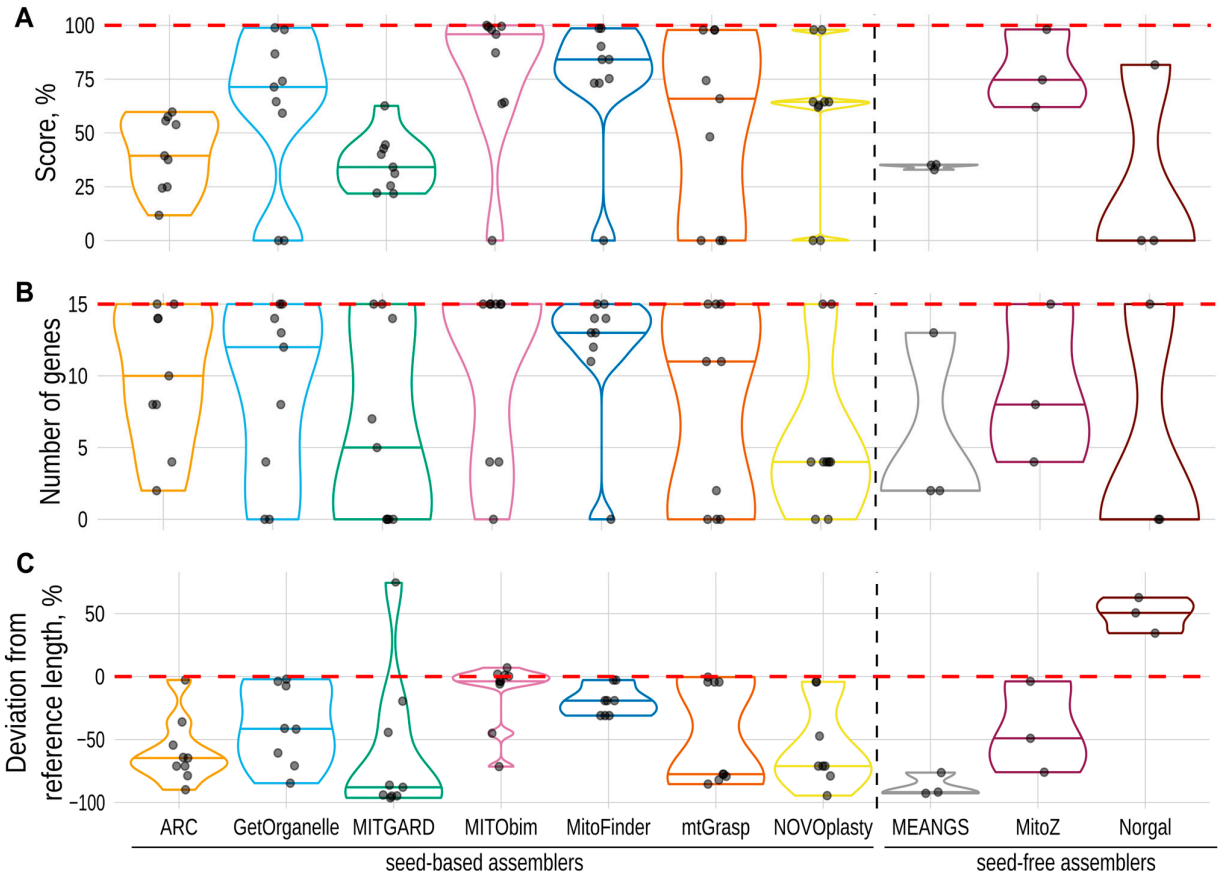


Figure 6. Overall statistics on assembly quality in the results produced by different assemblers on DNA sequencing data. (**A**) Assembly scores. (**B**) Number of found genes out of 13 protein-coding and 2 rRNA genes. (**C**) Relative deviation of the length of the obtained assembly from the reference assembly length. Each individual point represents a separate assembly attempt, and the violin plots show the overall distribution shape. Solid horizontal lines within violins correspond to median values. Dashed red horizontal lines mark the ideal target value of each parameter.

Our idea was therefore to combine two assemblers with very different approaches and strengths. MITObim had the best median score and median number of genes, but it depends heavily on the seed. MitoFinder was the second best in the number of genes; it uses a totally different strategy (de novo assembly) and also has the advantage of producing high-quality annotation. So, we suggest first using MitoFinder to obtain a good reference (even if incomplete) and then finishing it with MITObim if needed, and also annotating with MitoFinder again in this case (Figure 7A). Moreover, we found that k-mer-based prefiltering of reads based on the reference significantly lowered the computational load, allowing all computations to run on a laptop. This procedure requires a reasonably close reference, though it does not have to be of the same species or even genus.

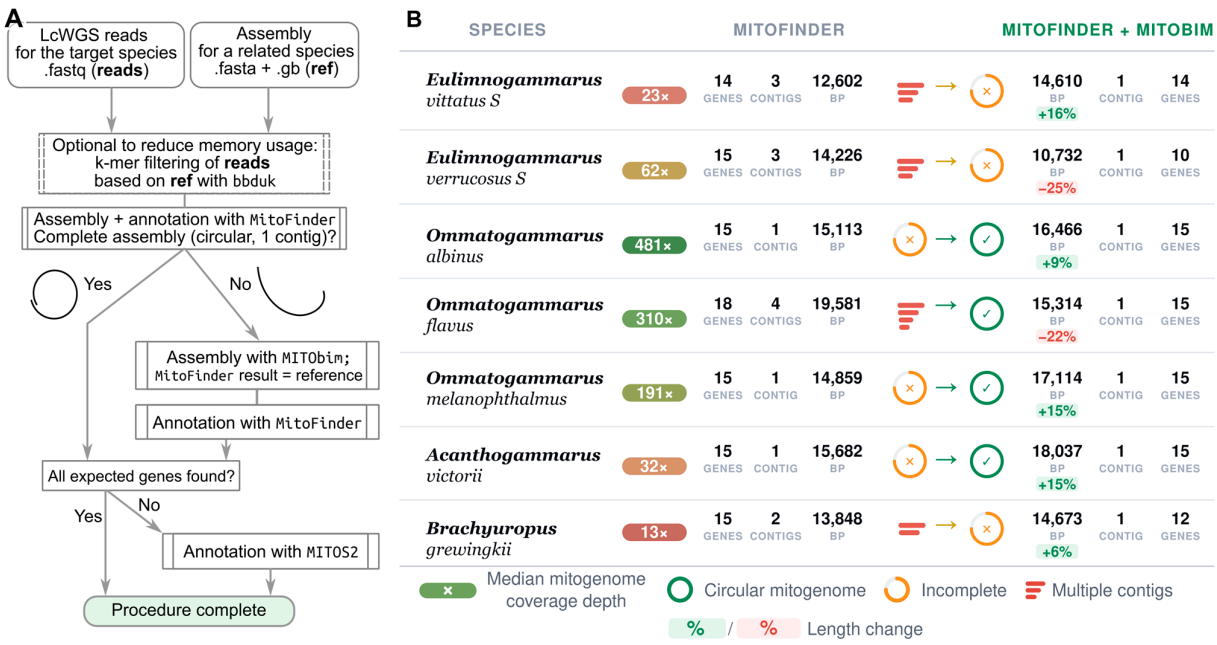


Figure 7. Recommended procedure for the assembly and annotation of a new mitochondrial genome. (**A**) Pipeline overview. Software titles are in monotype font. (**B**) Results of applying the recommended procedure to seven species of Baikal amphipods. See Table S7 for more details on assembly statistics.

To test the procedure, we applied it to reads from seven species of Baikal amphipods using the same RefSeq *E. cyaneus* assembly as a reference (Figure 7B). For two species belonging to the same genus, *Eulimnogammarus*, surprisingly, complete mitochondrial genomes were not assembled, even though MitoFinder recovered complete or almost complete sets of genes at the first step. It is important to note that the *E. vittatus S* sample was characterized by a relatively low median mitochondrial genome read depth of 23×, and in the original manuscript, the assembly was also incomplete and closed with Sanger sequencing [64]. In accordance with this, in the analysis of reads, an approximately 45 bp stretch of the published assembly was not covered by reads (Figure S3A). In the case of the *E. verrucosus S* reads, we additionally ran MITObim directly on the *E. cyaneus* reference in the hope of obtaining a circular assembly that can be used for aligning reads, but the result was not circular (Table S7). Thus, we decided to use *E. verrucosus W* as the closest reference for obtaining *E. verrucosus S* assembly. The resulting assembly was circular and had 15 predicted genes (Table S7). However, the alignment of reads showed multiple uncovered regions (Figure S3B), which, on closer inspection, turned out to be fragments of the *E. verrucosus W* reference. This *E. verrucosus* assembly was still used to estimate read depth, but it uncovered an unwanted behavior of MITObim.

For the three species of a related genus *Ommatogammarus*, the procedure worked well, as in the first step, incomplete mitogenomes with complete sets of genes were recovered with MitoFinder. In the case of *O. flavus*, three additional contigs with duplicated genes, which could be contamination or NUMT fragments, were found by MitoFinder. The longest of these contigs, containing duplicated *nad4* and *nad5* sequences, was ~97% identical to the corresponding part of the longest contig, which is overall consistent with its NUMT origin, while for the rest, no definitive conclusions could be made. However, in the second step, MITObim, run with default options, recovered complete circular mitochondrial genomes (Figure S3C–E).

Finally, two species that were slightly more distant from *Eulimnogammarus* [99] were analyzed. In the case of *A. victorii*, a complete circular genome was assembled using the suggested procedure, even though the original paper had provided an incomplete mitogenome [65]. To achieve this, the k for baiting reads in MITObim had to be lowered to k = 17, and the parameters for circularization in MitoFinder needed to be slightly relaxed as well (--circular-size 39 and --circular-offset 500), but the alignment of reads supported the circular structure in this as well as all the other assemblies (Figure S3). It is important to note that MitoFinder did not find the *atp8* gene in the *A. victorii* assembly, but an additional check with MITOS2 showed it was present. At the same time, MITOS2 missed *rrnL*, exactly as in the case of *E. cyaneus* (Figure 5), which means that the results from the two annotation pipelines may need to be combined. Surprisingly, in the case of *B. grewingkii*, for which a complete mitochondrial genome was published from the same data [63], our procedure failed to recover a complete genome. It is worth noting, however, that the median mitochondrial read depth for *A. victorii* is much higher than for *B. grewingkii* (calculated based on the published assembly): 32× and 13×, respectively.

4. Discussion

In this work, we compared the performance of ten tools for mitochondrial genome assembly and six tools for annotation using short-read genome and transcriptome sequencing data from several amphipod species. In this analysis, we kept in mind an inexperienced user, i.e., someone who can run pipelines but does not have time and/or the necessary experience to fine-tune and solve errors in the process and who may not have access to large computational power, so we considered no-installation access (Galaxy) and ways to reduce the computational load. We tried to resolve several questions.

(1) Which of the tools (if any) enable reconstruction of the full circular mitochondrial genome without manual finishing, at least for average mitochondrial genomes of ~16 kbp? Our sample of relatively hard-to-assemble genomes shows that multiple tools are capable of performing this task, depending on the input data (Figures 1 and 3). Ultimately, no single optimal solution exists, and multiple considerations should be taken into account. With RNA-seq, only MitoZ consistently provided high-quality assemblies, MITGARD always returned full gene sets, and MitoFinder and MITObim returned full gene sets in some cases (Figure 2). With lcWGS, multiple tools produced high-quality assemblies. In cases where the same-species or very close reference is available in the form of a complete or almost complete genome, MITObim might be a reasonable first choice (Figure 3), especially if computational resources are limited (Figure 4). However, it is very important to check the MITObim results for artifacts originating from the reference sequence (see Figure 3 and text below, as well as Figure S3B and related text). MitoFinder performs well with different kinds of input data (Figure 6) and also fulfills both tasks of assembly and annotation (see below). Moreover, four more tools, namely GetOrganelle, mtGrasp, NOVOplasty, and MitoZ, provided high-quality assemblies in particular cases (Figures 1 and 6) and should also be taken into account. Moreover, GetOrganelle, MITObim, and NOVOplasty have the advantage of being available from Galaxy (see Table 1), which alleviates the problems of finding computational resources and installing software. Finally, GetOrganelle is included in a number of automated pipelines allowing for automatic downloading of the data [36,37] and also works very fast compared with most other pipelines (Figure 4), and these characteristics make it suitable for large-scale projects connected to reanalysis of public data.

(2) Should one preferably perform genome skimming or transcriptome sequencing? According to our results, genome sequencing is preferable, as it provides more consistent results with comparable sequencing volumes when the main goal of the project is mitochondrial genome assembly. However, RNA-seq can still be used, especially if the main

goal is to obtain sets of mitochondrial genes, as demonstrated by multiple studies [16,18–20,58,170–172]. Only a few studies have assembled complete mitochondrial genomes from RNA-seq in different invertebrate taxa. They used the following tools: mitoRNA, a bowtie2- and Trinity-based script [20] that we failed to run, assembly of complete data with Trinity [58], and multiple rounds of mapping to a reference genome of a congener species with Geneious Prime v2020.2 [173].

(3) How much does the depth of sequencing coverage affect assembly quality, and what is the safe minimal depth for assembly? Here, we estimated the mitochondrial to nuclear read depth ratio for eight amphipod species and found that it varied approximately 9-fold, from 41 to 370, even in this limited sample, and mitochondrial genome read depth, quite predictably, turned out to correlate with assembly success (Figure 2). Moreover, the results of combining MitoFinder with MITObim (Figure 7) essentially corroborated this observation, as samples characterized by a median read depth <30× did not assemble into complete genomes. However, as seen with *E. verrucosus S*, high mitochondrial genome coverage does not guarantee a high-quality assembly (Figure 7B). Combined with the overall higher scores for mitogenome assemblies for *H. azteca*, which has a smaller nuclear genome, compared to those of *E. cyaneus* with a larger nuclear genome at the same mitochondrial read depth (Figure 2), this result suggests an important contribution of the nuclear genome contents, most probably elements such as NUMTs and repeated sequences, to the success of the mitochondrial genome assembly. Therefore, it is hard to make a universal recommendation for sequencing volume, but 1× depth of coverage for the nuclear genome can be suggested as a safe minimum for mitochondrial genome assembly for most species. Alternatively, it might be useful to estimate the mitochondrial-to-nuclear genome coverage depth ratio in a DNA sample before sequencing, for example, by comparing a mitochondrial sequence and a one-copy nuclear sequence with quantitative PCR, in order to more precisely estimate the necessary sequencing volumes, targeting >30× read depth of the mitochondrial genome.

(4) What can be used as a seed when a same-species reference is not available: a *cox1* fragment of the studied species or a mitochondrial genome of a related species? These options work with comparable success rates and should thus both be explored (Figure 3).

(5) What is the most accurate annotation tool? We found that MitoFinder is good at transferring annotation from a reference, but the reference needs to be closely related. MITOS2 should be the first choice in case a related genome is absent or the next logical step if the annotation produced by MitoFinder lacks some of the expected genes. This can happen with PCGs if the reference is too distant (see Section 3.3) or with tRNAs that span the zero coordinate if MitoFinder is provided with an assembly that is already trimmed to the circularization position (Figure 5). In these cases, MITOS2 can find missing PCGs because it uses a larger database and genes split by the zero coordinate because it takes into account circularization. It is also very important to carefully choose taxon-appropriate covariance models for tRNA genes.

Also, an important consideration before submitting a new mitochondrial genome to a database is standardization of the strand and start site, as fasta files are inherently linear, making it highly inconvenient to retrieve the sequences of those genes that are disrupted by the end/start. There are suggestions to standardize by starting from the tRNA-Phe gene (*trnF*) [26], but it raises several possible issues. For many assemblies, starting from *trnF* is possible but not convenient, as it is on the opposite chain relative to the majority of PCGs, and in this case *cox1*, being the major barcode sequence, would be reverse-complementary in the fasta sequence—thus, starting from the start codon of *cox1* is also very common (for example, [61,65,174]). The pancrustacean gene order is also commonly depicted starting from *cox1* [175]. Alternatively, circularMT depicts circular genomes as starting from *trnM* [176]. So, there are multiple alternative options, and researchers working with various

groups may have specific traditions. Moreover, standardization is either not feasible for incomplete mitogenomes, or it would require adding a roughly estimated or even arbitrary number of Ns. Thus, it is hard to provide a one-size-fits-all recommendation, but standardization can be important for studies with comparative analyses of multiple species.

Some benchmark studies comparing several tools have been published recently and provide valuable information, but they understandably do not keep up with the publication of all new tools. For example, there is a benchmark study on chloroplast genome assembly, which suggested GetOrganelle, Fast-Plast and NOVOPlasty as the most successful tools [136]. A benchmark devoted to insects, particularly Culicoides midges, compared six mappers, seven assemblers, and four annotation pipelines. It showed that mapping to a reference mitogenome generated better mitochondrial genomes than de novo assembly, but assembly with SPAdes, MEGAHIT, and MitoZ performed better if no closely related reference mitogenome was available, while the MITOS2 pipeline was the best-performing annotating tool [177]. A systematic comparison performed using human mitochondrial genomes as an example found that most tools tested adequately performed the task and that MToolBox, a specialized tool for the assembly of human mitochondrial genomes [178], showed the best performance, while NOVOPlasty was the second fastest and utilized the least computational resources when tested on simulated datasets [179]. Finally, a recent benchmark of 11 tools for mitochondrial plant genome assembly revealed five best-performing ones; four of them utilize long reads, and the remaining one was GetOrganelle [180]. Taken together, these and our results show that the results of benchmark studies are highly specific to the analyzed taxon, meaning that such analyses should be performed for particular groups at least at the beginning of large comparative studies. To make comparative analysis easier, we provide scripts for running multiple assemblers on the same dataset.

Potential weaknesses of our analysis include the fact that we did not test all possible combinations of parameters and tested on limited data. So, we cannot guarantee that our conclusions apply to any other group. However, our goal was to provide an overview of tools for this task and equip the interested reader with a methodology to perform the same comparison on their group of interest. Also, the suggested optimal procedure only allowed us to assemble complete circular mitogenomes in four out of seven cases and it mostly failed on datasets with low read depth.

5. Conclusions

A comparison of ten mitochondrial genome assembly tools and six annotation tools for three species of Baikal amphipods revealed no universal best solution. While MITObim and MitoFinder produced the best results for reference-based and de novo assembly of short-read DNA sequencing data, respectively, MitoZ was the best choice for reference-free assembly of short-read RNA sequencing data. MitoFinder was shown to be the best for transferring annotation from the reference and MITOS2 was the top option for de novo annotation, but its results require additional curation. Finally, we suggest a laptop-friendly procedure involving the combination of MITObim and MitoFinder to assemble and annotate a new mitochondrial genome from DNA sequencing data using a related reference, which has been tested on data for Baikal amphipods but could be reproduced for other taxa.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/d18120741/s1: Table S1: Amphipod mitochondrial genomes published in NCBI GenBank as of 31 August 2026; Table S2: Assembly tools and pipelines not used in this work; Table S3: tRNA annotation tools; Table S4: RNA annotation pipelines not used in this

work; Table S5: Gene coordinates obtained from different tools; Table S6: Comparative characteristic of mitochondrial genome assemblies obtained in this work; Table S7: Results of applying the recommended procedure to lcWGS for seven species of Baikal amphipods; Figure S1: Quality of mitochondrial genomes assembled with different seeds; Figure S2: Alignment of RNA-seq reads to the mitochondrial genome of *E. cyaneus* to confirm gene borders; Figure S3: Read alignment to the circular mitochondrial genome assemblies studied with the recommended procedure..

Author Contributions: Conceptualization, P.D.; methodology, A.M., I.O. and P.D.; validation, A.M., I.O. and P.D.; resources, M.T.; writing—original draft preparation, A.M. and P.D.; writing—review and editing, A.M., I.O., P.D. and M.T.; visualization, A.M., I.O. and P.D.; supervision, P.D.; project administration, P.D. and M.T.; funding acquisition, P.D. and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: The work was financed by the Russian Science Foundation, project number 22-14-00128-P (https://rscf.ru/project/22-14-00128/; accessed on 8 July 2026) and the Russian Ministry of Science and Higher Education, project number FZZE-2024-0008. Genome sequencing, assembly and annotation benchmarks of genome skimming and transcriptome sequencing data were supported by the Russian Science Foundation. Comparative analysis of annotation pipelines was supported by the Ministry of Science and Higher Education.

Institutional Review Board Statement: The animal study protocol was approved by the Animal Subjects Research Committee of the Institute of Biology at Irkutsk State University (protocol #9/2022; 18 October 2022).

Data Availability Statement: Data supporting the conclusions of this manuscript are available within the article and in the Supplementary Materials. Reproducible code is available at GitHub (https://github.com/drozdovapb/mt_genome_benchmark; accessed on 22 November 2026). The raw DNA sequencing reads obtained in this work are available from DDBJ/NCBI/ENA via accession numbers DRR911159, DRR911160, and DRR911165 (BioProject PRJDB40454). Newly assembled mitochondrial genomes are submitted to NCBI GenBank/DDBJ/ENA (accession numbers QB227943 for *B. pullus* and QB180845 for *E. cyaneus*) and are available at GitHub.

Acknowledgments: We would like to thank Ekaterina Govorukhina (Irkutsk State University) for her help with *B. pullus* identification and Daria Krasavina for her contribution to the early stages of benchmarking assemblers. We are also grateful to the team at the Institute of Lake Baikal Biology at Irkutsk State University for fruitful discussions and especially indebted to Anton Gurkov for his feedback on different versions of the manuscript and testing code reproducibility. The Galaxy server usegalaxy.eu used for some calculations is partly funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg. During the preparation of this manuscript, the authors used DeepSeek (version V4.1-Flash, accessed 9 October 2026) for the purposes of developing initial templates for Figure 7B and the graphical abstract. The authors have reviewed and extensively edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

bp	base pairs
kbp	kilobase pairs
Gbp	gigabase pairs

cox1, also COI or CO1	cytochrome c oxidase subunit I
CM	covariance model
HMM	hidden Markov model
lcWGS	low-coverage whole-genome sequencing (also genome skimming)
NGS	next-generation (short-read) sequencing
NUMT	nuclear mitochondrial DNA segment
pg	picogram
PCG	protein-coding gene
RNA-seq	RNA sequencing (also transcriptome sequencing)
TGS	third-generation (long-read) sequencing

References

1. Smith, D.R. The Past, Present and Future of Mitochondrial Genomics: Have We Sequenced Enough mtDNAs? *Brief. Funct. Genom.* **2016**, *15*, 47–54. https://doi.org/10.1093/bfgp/elv027.
2. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; deWaard, J.R. Biological Identifications through DNA Barcodes. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **2003**, *270*, 313–321. https://doi.org/10.1098/rspb.2002.2218.
3. Ratnasingham, S.; Wei, C.; Chan, D.; Agda, J.; Agda, J.; Ballesteros-Mejia, L.; Boutou, H.A.; El Bastami, Z.M.; Ma, E.; Manjunath, R.; et al. BOLD v4: A Centralized Bioinformatics Platform for DNA-Based Biodiversity Data. In *DNA Barcoding: Methods and Protocols*; DeSalle, R., Ed.; Springer: New York, NY, USA, 2024; pp. 403–441, ISBN 978-1-07-163581-0.
4. Moritz, C.; Dowling, T.E.; Brown, W.M. Evolution of Animal Mitochondrial DNA: Relevance for Population Biology and Systematics. *Annu. Rev. Ecol. Syst.* **1987**, *18*, 269–292.
5. Hubert, N.; Phillips, J.D.; Hanner, R.H. Delimiting Species with Single-Locus DNA Sequences. In *DNA Barcoding: Methods and Protocols*; DeSalle, R., Ed.; Springer: New York, NY, USA, 2024; pp. 53–76, ISBN 978-1-07-163581-0.
6. Zardoya, R. Recent Advances in Understanding Mitochondrial Genome Diversity. *F1000Research* **2020**, *9*, 270. https://doi.org/10.12688/f1000research.21490.1.
7. Kayal, E.; Lavrov, D.V. One Ring Does Not Rule Them All: Linear mtDNA in Metazoa. *Gene* **2025**, *933*, 148999. https://doi.org/10.1016/j.gene.2024.148999.
8. Lavrov, D.V.; 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.
9. Lavrov, D.V. Mitochondrial Genomes in Invertebrate Animals. In *Molecular Life Sciences*; Springer: New York, NY, USA, 2018; pp. 728–734, ISBN 978-1-4614-1531-2.
10. Ghiselli, F.; Gomes-dos-Santos, A.; Adema, C.M.; Lopes-Lima, M.; Sharbrough, J.; Boore, J.L. Molluscan Mitochondrial Genomes Break the Rules. *Phil. Trans. R. Soc. B* **2021**, *376*, 20200159. https://doi.org/10.1098/rstb.2020.0159.
11. Briscoe, A.G.; Goodacre, S.; Masta, S.E.; Taylor, M.I.; Arnedo, M.A.; Penney, D.; Kenny, J.; Creer, S. Can Long-Range PCR Be Used to Amplify Genetically Divergent Mitochondrial Genomes for Comparative Phylogenetics? A Case Study Within Spiders (Arthropoda: Araneae). *PLoS ONE* **2013**, *8*, e62404. https://doi.org/10.1371/journal.pone.0062404.
12. Zascavage, R.R.; Hall, C.L.; Thorson, K.; Mahmoud, M.; Sedlazeck, F.J.; Planz, J.V. Approaches to Whole Mitochondrial Genome Sequencing on the Oxford Nanopore MinION. *Curr. Protoc. Hum. Genet.* **2019**, *104*, e94. https://doi.org/10.1002/cphg.94.
13. Baeza, J.A. Yes, We Can Use It: A Formal Test on the Accuracy of Low-Pass Nanopore Long-Read Sequencing for Mitophylogenomics and Barcoding Research Using the Caribbean Spiny Lobster *Panulirus argus*. *BMC Genom.* **2020**, *21*, 882. https://doi.org/10.1186/s12864-020-07292-5.
14. 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 Genom.* **2020**, *21*, 785. https://doi.org/10.1186/s12864-020-07183-9.
15. Gan, H.M.; Linton, S.M.; Austin, C.M. Two Reads to Rule Them All: Nanopore Long Read-Guided Assembly of the Iconic Christmas Island Red Crab, *Gecarcoidea natalis* (Pocock, 1888), Mitochondrial Genome and the Challenges of AT-Rich Mitogenomes. *Mar. Genom.* **2019**, *45*, 64–71. https://doi.org/10.1016/j.margen.2019.02.002.
16. Allio, R.; Schomaker-Bastos, A.; Romiguier, J.; Prosdocimi, F.; Nabholz, B.; Delsuc, F. MitoFinder: Efficient Automated Large-scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics. *Mol. Ecol. Resour.* **2020**, *20*, 892–905. https://doi.org/10.1111/1755-0998.13160.
17. Raxworthy, C.J.; Smith, B.T. Mining Museums for Historical DNA: Advances and Challenges in Museomics. *Trends Ecol. Evol.* **2021**, *36*, 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009.

18. Kobayashi, G. Buried Treasure in a Public Repository: Mining Mitochondrial Genes of 32 Annelid Species from Sequence Reads Deposited in the Sequence Read Archive (SRA). *PeerJ* **2023**, *11*, e16446. https://doi.org/10.7717/peerj.16446.
19. Smith, D.R. RNA-Seq Data: A Goldmine for Organelle Research. *Brief. Funct. Genom.* **2013**, *12*, 454–456. https://doi.org/10.1093/bfgp/els066.
20. Forni, G.; Puccio, G.; Bourguignon, T.; Evans, T.; Mantovani, B.; Rota-Stabelli, O.; Luchetti, A. Complete Mitochondrial Genomes from Transcriptomes: Assessing Pros and Cons of Data Mining for Assembling New Mitogenomes. *Sci. Rep.* **2019**, *9*, 14806. https://doi.org/10.1038/s41598-019-51313-7.
21. Lozano-Fernandez, J. A Practical Guide to Design and Assess a Phylogenomic Study. *Genome Biol. Evol.* **2022**, *14*, evac129. https://doi.org/10.1093/gbe/evac129.
22. Chalifa-Caspi, V. RNA-Seq in Nonmodel Organisms. In *Deep Sequencing Data Analysis*; Shomron, N., Ed.; Springer: New York, NY, USA, 2021; pp. 143–167, ISBN 978-1-07-161103-6.
23. Sanita Lima, M.; Silva Domingues, D.; Rossi Paschoal, A.; Smith, D.R. Long-Read RNA Sequencing Can Probe Organelle Genome Pervasive Transcription. *Brief. Funct. Genom.* **2024**, *23*, 695–701. https://doi.org/10.1093/bfgp/elae026.
24. Hahn, C.; Bachmann, L.; Chevreux, B. Reconstructing Mitochondrial Genomes Directly from Genomic Next-Generation Sequencing Reads—A Baiting and Iterative Mapping Approach. *Nucleic Acids Res.* **2013**, *41*, e129. https://doi.org/10.1093/nar/gkt371.
25. Nachtigall, P.G.; Grazziotin, F.G.; Junqueira-de-Azevedo, I.L.M. MITGARD: An Automated Pipeline for Mitochondrial Genome Assembly in Eukaryotic Species Using RNA-Seq Data. *Brief. Bioinform.* **2021**, *22*, bbaa429. https://doi.org/10.1093/bib/bbaa429.
26. Lopez, M.L.D.; Yang, C.L.; Coombe, L.; Warren, R.L.; Allison, M.J.; Imbery, J.J.; Birol, I.; Helbing, C.C. mtGrasp: Streamlined Reference-Grade Mitochondrial Genome Assembly and Standardization to Enhance Metazoan Mitogenome Resources. *Methods Ecol. Evol.* **2025**, *16*, 668–677. https://doi.org/10.1111/2041-210X.14506.
27. Song, M.-H.; Yan, C.; Li, J.-T. MEANGS: An Efficient Seed-Free Tool for de Novo Assembling Animal Mitochondrial Genome Using Whole Genome NGS Data. *Brief. Bioinform.* **2022**, *23*, bbab538. https://doi.org/10.1093/bib/bbab538.
28. Jin, J.-J.; Yu, W.-B.; Yang, J.-B.; Song, Y.; dePamphilis, C.W.; Yi, T.-S.; Li, D.-Z. GetOrganelle: A Fast and Versatile Toolkit for Accurate de Novo Assembly of Organelle Genomes. *Genome Biol.* **2020**, *21*, 241. https://doi.org/10.1186/s13059-020-02154-5.
29. Al-Nakeeb, K.; Petersen, T.N.; Sicheritz-Pontén, T. Norgal: Extraction and de Novo Assembly of Mitochondrial DNA from Whole-Genome Sequencing Data. *BMC Bioinform.* **2017**, *18*, 510. https://doi.org/10.1186/s12859-017-1927-y.
30. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data. *Nucleic Acids Res.* **2017**, *45*, e18. https://doi.org/10.1093/nar/gkw955.
31. Meng, G.; Li, Y.; Yang, C.; Liu, S. MitoZ: A Toolkit for Animal Mitochondrial Genome Assembly, Annotation and Visualization. *Nucleic Acids Res.* **2019**, *47*, e63. https://doi.org/10.1093/nar/gkz173.
32. Hunter, S.S.; Lyon, R.T.; Sarver, B.A.J.; Hardwick, K.; Forney, L.J.; Settles, M.L. Assembly by Reduced Complexity (ARC): A Hybrid Approach for Targeted Assembly of Homologous Sequences. *bioRxiv* **2015**, https://doi.org/10.1101/014662.
33. Li, J.-Y.; Li, W.-X.; Wang, A.-T.; Zhang, Y. MitoFlex: An Efficient, High-Performance Toolkit for Animal Mitogenome Assembly, Annotation and Visualization. *Bioinformatics* **2021**, *37*, 3001–3003. https://doi.org/10.1093/bioinformatics/btab111.
34. Schomaker-Bastos, A.; Prosdocimi, F. mitoMaker: A Pipeline for Automatic Assembly and Annotation of Animal Mitochondria Using Raw NGS Data. *Preprints* **2018**, https://doi.org/10.20944/preprints201808.0423.v1.
35. Lubośny, M.; Przyłucka, A.; Śmietanka, B.; Burzyński, A. *Semimytilus algosus*: First Known Hermaphroditic Mussel with Doubly Uniparental Inheritance of Mitochondrial DNA. *Sci. Rep.* **2020**, *10*, 11256. https://doi.org/10.1038/s41598-020-67976-6.
36. Eldem, V.; Balcı, M.A. Mining NCBI Sequence Read Archive Database: An Untapped Source of Organelle Genomes for Taxonomic and Comparative Genomics Research. *Diversity* **2024**, *16*, 104. https://doi.org/10.3390/d16020104.
37. White, O.W.; Hall, A.; Price, B.W.; Williams, S.T.; Clark, M.D. A Snakemake Toolkit for the Batch Assembly, Annotation and Phylogenetic Analysis of Mitochondrial Genomes and Ribosomal Genes from Genome Skims of Museum Collections. *Mol. Ecol. Resour.* **2025**, *25*, e14036. https://doi.org/10.1111/1755-0998.14036.
38. Ozerova, I.; Fallmann, J.; Mörl, M.; Bernt, M.; Prohaska, S.J.; Stadler, P.F. Aberrant Mitochondrial tRNA Genes Appear Frequently in Animal Evolution. *Genome Biol. Evol.* **2024**, *16*, evae232. https://doi.org/10.1093/gbe/evae232.
39. Doublet, V.; Ubrig, E.; Alioua, A.; Bouchon, D.; Marcadé, I.; Maréchal-Drouard, L. Large Gene Overlaps and tRNA Processing in the Compact Mitochondrial Genome of the Crustacean *Armadillidium vulgare*. *RNA Biol.* **2015**, *12*, 1159–1168. https://doi.org/10.1080/15476286.2015.1090078.
40. Rawlings, T.A.; Collins, T.M.; Bieler, R. Changing Identities: tRNA Duplication and Remolding within Animal Mitochondrial Genomes. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15700–15705. https://doi.org/10.1073/pnas.2535036100.

41. Sahyoun, A.H.; Hölzer, M.; Jühling, F.; Höner zu Siederdissen, C.; Al-Arab, M.; Tout, K.; Marz, M.; Middendorf, M.; Stadler, P.F.; Bernt, M. Towards a Comprehensive Picture of Alloacceptor tRNA Remolding in Metazoan Mitochondrial Genomes. *Nucleic Acids Res.* **2015**, *43*, 8044–8056. https://doi.org/10.1093/nar/gkv746.
42. Clayton, D. Mitochondrial DNA Replication: What We Know. *IUBMB Life* **2003**, *55*, 213–217. https://doi.org/10.1080/1521654031000134824.
43. Liu, Y.; Liu, H.; Zhang, F.; Xu, H. The Initiation of Mitochondrial DNA Replication. *Biochem. Soc. Trans.* **2024**, *52*, 1243–1251. https://doi.org/10.1042/BST20230952.
44. Cameron, S.L. How to Sequence and Annotate Insect Mitochondrial Genomes for Systematic and Comparative Genomics Research. *Syst. Entomol.* **2014**, *39*, 400–411. https://doi.org/10.1111/syen.12071.
45. Jakovlić, I.; Zou, H.; Zhao, X.-M.; Zhang, J.; Wang, G.-T.; Zhang, D. Evolutionary History of Inversions in Directional Mutational Pressures in Crustacean Mitochondrial Genomes: Implications for Evolutionary Studies. *Mol. Phylogenet. Evol.* **2021**, *164*, 107288. https://doi.org/10.1016/j.ympev.2021.107288.
46. Nicholls, T.J.; Minczuk, M. In D-Loop: 40 Years of Mitochondrial 7S DNA. *Exp. Gerontol.* **2014**, *56*, 175–181. https://doi.org/10.1016/j.exger.2014.03.027.
47. Arfianti, T.; Wilson, S.; Costello, M.J. Progress in the Discovery of Amphipod Crustaceans. *PeerJ* **2018**, *6*, e5187. https://doi.org/10.7717/peerj.5187.
48. Cristescu, M.E.; Adamowicz, S.J.; Vaillant, J.J.; Haffner, D.G. Ancient Lakes Revisited: From the Ecology to the Genetics of Speciation. *Mol. Ecol.* **2010**, *19*, 4837–4851. https://doi.org/10.1111/j.1365-294X.2010.04832.x.
49. Hultgren, K.M.; Jeffery, N.W.; Moran, A.; Gregory, T.R. Latitudinal Variation in Genome Size in Crustaceans. *Biol. J. Linn. Soc.* **2018**, *123*, 348–359. https://doi.org/10.1093/biolinnean/blx153.
50. Gregory, T.R. Animal Genome Size Database. Available online: https://www.genomesize.com/ (accessed on 29 July 2026).
51. Doležel, J.; Bartoš, J.; Voglmayr, H.; Greilhuber, J. Letter to the Editor. *Cytom. Part A* **2003**, *51A*, 127–128. https://doi.org/10.1002/cyto.a.10013.
52. Sterling-Montealegre, R.A.; Prada, C.F. Variability and Evolution of Gene Order Rearrangement in Mitochondrial Genomes of Arthropods (except Hexapoda). *Gene* **2024**, *892*, 147906. https://doi.org/10.1016/j.gene.2023.147906.
53. Romanova, E.V.; Bukin, Y.S.; Mikhailov, K.V.; Logacheva, M.D.; Aleoshin, V.V.; Sherbakov, D.Y. Hidden Cases of tRNA Gene Duplication and Remolding in Mitochondrial Genomes of Amphipods. *Mol. Phylogenet. Evol.* **2020**, *144*, 106710. https://doi.org/10.1016/j.ympev.2019.106710.
54. Kilpert, F.; Podsiadlowski, L. The Mitochondrial Genome of the Japanese Skeleton Shrimp *Caprella mutica* (Amphipoda: Caprellidea) Reveals a Unique Gene Order and Shared Apomorphic Translocations with Gammaridea. *Mitochondrial DNA* **2010**, *21*, 77–86. https://doi.org/10.3109/19401736.2010.490832.
55. Cook, C.E.; Yue, Q.; Akam, M. Mitochondrial Genomes Suggest That Hexapods and Crustaceans Are Mutually Paraphyletic. *Proc. R. Soc. B* **2005**, *272*, 1295–1304. https://doi.org/10.1098/rspb.2004.3042.
56. Ito, A.; Aoki, M.N.; Yokobori, S.; Wada, H. The Complete Mitochondrial Genome of *Caprella scaura* (Crustacea, Amphipoda, Caprellidea), with Emphasis on the Unique Gene Order Pattern and Duplicated Control Region. *Mitochondrial DNA* **2010**, *21*, 183–190. https://doi.org/10.3109/19401736.2010.517834.
57. Bauzà-Ribot, M.M.; Jaume, D.; Juan, C.; Pons, J. The Complete Mitochondrial Genome of the Subterranean Crustacean *Metacrangonyx longipes* (Amphipoda): A Unique Gene Order and Extremely Short Control Region. *Mitochondrial DNA* **2009**, *20*, 88–99. https://doi.org/10.1080/19401730902964417.
58. Mamos, T.; Grabowski, M.; Rewicz, T.; Bojko, J.; Strapagiel, D.; Burzyński, A. Mitochondrial Genomes, Phylogenetic Associations, and SNP Recovery for the Key Invasive Ponto-Caspian Amphipods in Europe. *Int. J. Mol. Sci.* **2021**, *22*, 10300. https://doi.org/10.3390/ijms221910300.
59. Bauzà-Ribot, M.M.; Juan, C.; Nardi, F.; Oromí, P.; Pons, J.; Jaume, D. Mitogenomic Phylogenetic Analysis Supports Continental-Scale Vicariance in Subterranean Thalassoid Crustaceans. *Curr. Biol.* **2012**, *22*, 2069–2074. https://doi.org/10.1016/j.cub.2012.09.012.
60. Ki, J.-S.; Hop, H.; Kim, S.-J.; Kim, I.-C.; Park, H.G.; Lee, J.-S. Complete Mitochondrial Genome Sequence of the Arctic Gammarid, *Onisimus nanseni* (Crustacea; Amphipoda): Novel Gene Structures and Unusual Control Region Features. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2010**, *5*, 105–115. https://doi.org/10.1016/j.cbd.2010.02.002.
61. Krebes, L.; Bastrop, R. The Mitogenome of *Gammarus duebeni* (Crustacea Amphipoda): A New Gene Order and Non-Neutral Sequence Evolution of Tandem Repeats in the Control Region. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2012**, *7*, 201–211. https://doi.org/10.1016/j.cbd.2012.02.004.

62. Shin, S.C.; Cho, J.; Lee, J.K.; Ahn, D.H.; Lee, H.; Park, H. Complete Mitochondrial Genome of the Antarctic Amphipod *Gondogeneia antarctica* (Crustacea, Amphipod). *Mitochondrial DNA* **2012**, *23*, 25–27. https://doi.org/10.3109/19401736.2011.643877.
63. Romanova, E.V.; Mikhailov, K.V.; Logacheva, M.D.; Kamaltynov, R.M.; Aleoshin, V.V.; Sherbakov, D.Y. The Complete Mitochondrial Genome of a Deep-Water Baikalian Amphipoda *Brachyuropus grewingkii* (Dybowsky, 1874). *Mitochondrial DNA Part A* **2016**, *27*, 4158–4159. https://doi.org/10.3109/19401736.2014.1003891.
64. Romanova, E.V.; Mikhailov, K.V.; Logacheva, M.D.; Kamaltynov, R.M.; Aleoshin, V.V.; Sherbakov, D.Y. The Complete Mitochondrial Genome of Baikalian Amphipoda *Eulimnogammarus vittatus* Dybowsky, 1874. *Mitochondrial DNA Part A* **2016**, *27*, 1795–1797. https://doi.org/10.3109/19401736.2014.963817.
65. Romanova, E.V.; Aleoshin, V.V.; Kamaltynov, R.M.; Mikhailov, K.V.; Logacheva, M.D.; Sirotinina, E.A.; Gornov, A.Y.; Anikin, A.S.; Sherbakov, D.Y. Evolution of Mitochondrial Genomes in Baikalian Amphipods. *BMC Genom.* **2016**, *17*, 1016. https://doi.org/10.1186/s12864-016-3357-z.
66. Lan, Y.; Sun, J.; Bartlett, D.H.; Rouse, G.W.; Tabata, H.G.; Qian, P.-Y. The Deepest Mitochondrial Genome Sequenced from Mariana Trench *Hirondellea gigas* (Amphipoda). *Mitochondrial DNA Part B* **2016**, *1*, 802–803. https://doi.org/10.1080/23802359.2016.1214549.
67. Aunins, A.W.; Nelms, D.L.; Hobson, C.S.; King, T.L. Comparative Mitogenomic Analyses of Three North American Stygobiont Amphipods of the Genus *Stygobromus* (Crustacea: Amphipoda). *Mitochondrial DNA Part B* **2016**, *1*, 560–563. https://doi.org/10.1080/23802359.2016.1174086.
68. Macher, J.N.; Leese, F.; Weigand, A.M.; Rozenberg, A. The Complete Mitochondrial Genome of a Cryptic Amphipod Species from the *Gammarus fossarum* Complex. *Mitochondrial DNA Part B* **2017**, *2*, 17–18. https://doi.org/10.1080/23802359.2016.1275844.
69. Hiki, K.; Ariyama, H.; Nakajima, N. The Complete Mitochondrial Genomes of Two Amphipod Species of the Genus *Grandidierella* (Crustacea: Amphipoda). *Mitochondrial DNA Part B* **2020**, *5*, 1535–1536. https://doi.org/10.1080/23802359.2020.1742215.
70. Hiki, K.; Ariyama, H.; Nakajima, N. The Complete Mitochondrial Genome of the Estuarine Amphipod *Grandidierella osakaensis* (Crustacea: Amphipoda). *Mitochondrial DNA Part B* **2020**, *5*, 3305–3306. https://doi.org/10.1080/23802359.2020.1778559.
71. Stokkan, M.; Jurado-Rivera, J.A.; Juan, C.; Jaume, D.; Pons, J. Mitochondrial Genome Rearrangements at Low Taxonomic Levels: Three Distinct Mitogenome Gene Orders in the Genus *Pseudoniphargus* (Crustacea: Amphipoda). *Mitochondrial DNA Part A* **2016**, *27*, 3579–3589. https://doi.org/10.3109/19401736.2015.1079821.
72. Juan, C.; Jurado-Rivera, J.A.; Moreno, E.; Wolff, C.; Jaume, D.; Pons, J. The Mitogenome of the Amphipod *Hyalella lucifugax* (Crustacea) and Its Phylogenetic Placement. *Mitochondrial DNA Part B* **2016**, *1*, 755–756. https://doi.org/10.1080/23802359.2016.1214554.
73. Beermann, J.; Westbury, M.V.; Hofreiter, M.; Hilgers, L.; Deister, F.; Neumann, H.; Raupach, M.J. Cryptic Species in a Well-Known Habitat: Applying Taxonomics to the Amphipod Genus *Epimeria* (Crustacea, Peracarida). *Sci. Rep.* **2018**, *8*, 6893. https://doi.org/10.1038/s41598-018-25225-x.
74. Yang, H.-M.; Song, J.-H.; Kim, M.-S.; Min, G.-S. The Complete Mitochondrial Genomes of Two Talitrid Amphipods, *Platorchestia japonica* and *P. parapacifica* (Crustacea, Amphipoda). *Mitochondrial DNA Part B* **2017**, *2*, 757–758. https://doi.org/10.1080/23802359.2017.1398606.
75. Cormier, A.; Wattier, R.; Teixeira, M.; Rigaud, T.; Cordaux, R. The Complete Mitochondrial Genome of *Gammarus roeselii* (Crustacea, Amphipoda): Insights into Mitogenome Plasticity and Evolution. *Hydrobiologia* **2018**, *825*, 197–210. https://doi.org/10.1007/s10750-018-3578-z.
76. Lee, C.-W.; Nakano, T.; Tomikawa, K.; Min, G.-S. The Complete Mitochondrial Genome of *Pseudocrangonyx daejeonensis* (Crustacea: Amphipoda: Pseudocrangonyctidae). *Mitochondrial DNA Part B* **2018**, *3*, 823–824. https://doi.org/10.1080/23802359.2018.1495116.
77. Li, J.; Zeng, C.; Yan, G.; He, L. Characterization of the Mitochondrial Genome of an Ancient Amphipod *Halice* sp. MT-2017 (Pardaliscidae) from 10,908 m in the Mariana Trench. *Sci. Rep.* **2019**, *9*, 2610. https://doi.org/10.1038/s41598-019-38735-z.
78. Kumar Patra, A.; Kim, M.-S.; Jung, T.W.; Cho, I.-Y.; Yoon, M.; Choi, J.-H.; Yang, Y. The Complete Mitochondrial Genome of the Sand-Hopper *Trinorchestia longiramus* (Amphipoda: Talitridae). *Mitochondrial DNA Part B* **2019**, *4*, 2104–2105. https://doi.org/10.1080/23802359.2019.1623100.
79. Stokkan, M.; Jurado-Rivera, J.A.; Oromí, P.; Juan, C.; Jaume, D.; Pons, J. Species Delimitation and Mitogenome Phylogenetics in the Subterranean Genus *Pseudoniphargus* (Crustacea: Amphipoda). *Mol. Phylogenet. Evol.* **2018**, *127*, 988–999. https://doi.org/10.1016/j.ympev.2018.07.002.
80. Li, J.; Song, Z.; Yan, G.; He, L. The Complete Mitochondrial Genome of the Largest Amphipod, *Alicella gigantea*: Insight into Its Phylogenetic Relationships and Deep Sea Adaptive Characters. *Int. J. Biol. Macromol.* **2019**, *141*, 570–577. https://doi.org/10.1016/j.ijbiomac.2019.09.050.
81. Lee, S.-H.; Lee, S.-H.; Lim, B.-J.; Kim, M.K.; Shin, M.-H. The Complete Mitochondrial Genome of *Ampithoe lacertosa* Spence Bate, 1858 (Crustacea: Amphipoda: Ampithoidae). *Mitochondrial DNA Part B* **2019**, *4*, 750–751. https://doi.org/10.1080/23802359.2019.1565965.

82. Lee, S.-H.; Wongkamhaeng, K.; Lee, S.-H.; Shin, M.-H. The Complete Mitochondrial Genome of *Pleonexes koreana* (Kim & Kim, 1988) (Crustacea: Amphipoda: Ampithoidae). *Mitochondrial DNA Part B* **2019**, *4*, 787–788. https://doi.org/10.1080/23802359.2019.1566793.
83. Bojko, J. The Mitochondrial Genome of UK (Non-Native) *Dikerogammarus haemobaphes* (Amphipoda: Gammaridae) Informs upon *Dikerogammarus* Evolution, Invasions and Associated Microparasites. *Hydrobiologia* **2020**, *847*, 229–242. https://doi.org/10.1007/s10750-019-04084-1.
84. Pons, J.; Jurado-Rivera, J.A.; Jaume, D.; Vonk, R.; Bauzà-Ribot, M.M.; Juan, C. The Age and Diversification of Metacrangonyctid Subterranean Amphipod Crustaceans Revisited. *Mol. Phylogenet. Evol.* **2019**, *140*, 106599. https://doi.org/10.1016/j.ympev.2019.106599.
85. Benito, J.B.; Porter, M.L.; Niemiller, M.L. The Mitochondrial Genomes of Five Spring and Groundwater Amphipods of the Family Crangonyctidae (Crustacea: Amphipoda) from Eastern North America. *Mitochondrial DNA Part B* **2021**, *6*, 1662–1667. https://doi.org/10.1080/23802359.2021.1926350.
86. Sun, S.; Wu, Y.; Ge, X.; Jakovlić, I.; Zhu, J.; Mahboob, S.; Al-Ghanim, K.A.; Al-Misned, F.; Fu, H. Disentangling the Interplay of Positive and Negative Selection Forces That Shaped Mitochondrial Genomes of *Gammarus pisinnus* and *Gammarus lacustris*. *R. Soc. Open Sci.* **2020**, *7*, 190669. https://doi.org/10.1098/rsos.190669.
87. Li, J.; Liao, Y.; Li, J.; He, L. The Complete Mitochondrial Genome of the Deep-Sea Amphipod *Eurythenes magellanicus* (Crustacea: Amphipoda: Lysianassidae). *Mitochondrial DNA Part B* **2020**, *5*, 337–339. https://doi.org/10.1080/23802359.2019.1703573.
88. Romanova, E.V.; Bukin, Y.S.; Mikhailov, K.V.; Logacheva, M.D.; Aleoshin, V.V.; Sherbakov, D.Y. The Mitochondrial Genome of a Freshwater Pelagic Amphipod *Macrohectopus branickii* Is among the Longest in Metazoa. *Genes* **2021**, *12*, 2030. https://doi.org/10.3390/genes12122030.
89. Lee, C.-W.; Nakano, T.; Tomikawa, K.; Min, G.-S. Complete Mitochondrial Genome of *Pseudocrangonyx joolaei* (Crustacea: Amphipoda: Pseudocrangonyctidae). *Mitochondrial DNA Part B* **2020**, *5*, 3011–3012. https://doi.org/10.1080/23802359.2020.1797592.
90. Lee, C.-W.; Tomikawa, K.; Min, G.-S. The Complete Mitochondrial Genome of *Jesogammarus* (*Jesogammarus*) *hinumensis* (Crustacea: Amphipoda: Anisogammaridae). *Mitochondrial DNA Part B* **2020**, *5*, 2436–2437. https://doi.org/10.1080/23802359.2020.1775520.
91. Zapelloni, F.; Pons, J.; Jurado-Rivera, J.A.; Jaume, D.; Juan, C. Phylogenomics of the *Hyalella* Amphipod Species-Flock of the Andean Altiplano. *Sci. Rep.* **2021**, *11*, 366. https://doi.org/10.1038/s41598-020-79620-4.
92. Yin, J.; Zhao, M.; Zhang, Z.; Fu, Y.; Liu, Y.; Liu, X.; Xu, J.; Gui, Z.; Tang, S.; Zhang, T. The Complete Mitogenome of Freshwater Gammarid *Grandidierella taihuensis* (Crustacea: Amphipoda). *Mitochondrial DNA Part B* **2021**, *6*, 1369–1371. https://doi.org/10.1080/23802359.2021.1910079.
93. Salabao, L.; Plevoets, T.; Frédérich, B.; Lepoint, G.; Kochzius, M.; Schön, I. Describing Novel Mitochondrial Genomes of Antarctic Amphipods. *Mitochondrial DNA Part B* **2022**, *7*, 810–818. https://doi.org/10.1080/23802359.2022.2073837.
94. Rabinowitz, C.N.; Brown, S.D.; McAllister, S.M.; Winans, A.K.; Keister, J.E.; Galaska, M.P. The Complete Mitochondrial Genome of *Cyphocaris challengeri* (Amphipoda: Cyphocarididae). *Mitochondrial DNA Part B* **2023**, *8*, 1128–1131. https://doi.org/10.1080/23802359.2023.2270206.
95. de la Cruz-Huervana, J.J.; Miyamoto, N.; Kano, Y.; Onikura, N.; Kurita, Y. The Complete Mitochondrial Genome of Freshwater Gammarid *Gammarus nipponensis* (Crustacea: Amphipoda: Gammaridae). *Mitochondrial DNA Part B* **2024**, *9*, 447–451. https://doi.org/10.1080/23802359.2024.2335990.
96. Macher, J.-N.; Šidagytė-Copilas, E.; Copilaș-Ciocianu, D. Comparative Mitogenomics of Native European and Alien Ponto-Caspian Amphipods. *NeoBiota* **2023**, *87*, 27–44. https://doi.org/10.3897/neobiota.87.105941.
97. Sherbakov, D.Y. Molecular Phylogenetic Studies on the Origin of Biodiversity in Lake Baikal. *Trends Ecol. Evol.* **1999**, *14*, 92–95. https://doi.org/10.1016/S0169-5347(98)01543-2.
98. Macdonald, K.S., III; Yampolsky, L.; Duffy, J.E. Molecular and Morphological Evolution of the Amphipod Radiation of Lake Baikal. *Mol. Phylogenet. Evol.* **2005**, *35*, 323–343. https://doi.org/10.1016/j.ympev.2005.01.013.
99. Naumenko, S.A.; Logacheva, M.D.; Popova, N.V.; Klepikova, A.V.; Penin, A.A.; Bazykin, G.A.; Etingova, A.E.; Mugue, N.S.; Kondrashov, A.S.; Yampolsky, L.Y. Transcriptome-Based Phylogeny of Endemic Lake Baikal Amphipod Species Flock: Fast Speciation Accompanied by Frequent Episodes of Positive Selection. *Mol. Ecol.* **2017**, *26*, 536–553. https://doi.org/10.1111/mec.13927.
100. Drozdova, P.B.; Madyarova, E.V.; Gurkov, A.N.; Saranchina, A.E.; Romanova, E.V.; Petunina, J.V.; Peretolchina, T.E.; Sherbakov, D.Y.; Timofeyev, M.A. Lake Baikal Amphipods and Their Genomes, Great and Small. *Vestn. VOGiS* **2024**, *28*, 317–325. https://doi.org/10.18699/vjgb-24-36.
101. Jeffery, N.W.; Yampolsky, L.; Gregory, T.R. Nuclear DNA Content Correlates with Depth, Body Size, and Diversification Rate in Amphipod Crustaceans from Ancient Lake Baikal, Russia. *Genome* **2017**, *60*, 303–309. https://doi.org/10.1139/gen-2016-0128.

102. Yuxiang, W.; Peretolchina, T.E.; Romanova, E.V.; Sherbakov, D.Y. Comparison of the Evolutionary Patterns of DNA Repeats in Ancient and Young Invertebrate Species Flocks of Lake Baikal. *Vavilovskii Zhurnal Genet. Selektsii* **2023**, *27*, 349–356. https://doi.org/10.18699/VJGB-23-42.
103. Rivarola-Duarte, L.; Otto, C.; Jühling, F.; Schreiber, S.; Bedulina, D.; Jakob, L.; Gurkov, A.; Axenov-Gribanov, D.; Sahyoun, A.H.; Lucassen, M.; et al. A First Glimpse at the Genome of the Baikalian Amphipod *Eulimnogammarus verrucosus*. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2014**, *322*, 177–189. https://doi.org/10.1002/jez.b.22560.
104. Fišer, C.; Robinson, C.T.; Malard, F. Cryptic Species as a Window into the Paradigm Shift of the Species Concept. *Mol. Ecol.* **2018**, *27*, 613–635. https://doi.org/10.1111/mec.14486.
105. Jourdan, J.; Bundschuh, M.; Copilaș-Ciocianu, D.; Fišer, C.; Grabowski, M.; Hupało, K.; Jemec Kokalj, A.; Kabus, J.; Römbke, J.; Soose, L.J.; et al. Cryptic Species in Ecotoxicology. *Environ. Toxicol. Chem.* **2023**, *42*, 1889–1914. https://doi.org/10.1002/etc.5696.
106. Copilaș-Ciocianu, D.; Sidorov, D. Taxonomic, Ecological and Morphological Diversity of Ponto-Caspian Gammaroidean Amphipods: A Review. *Org. Divers. Evol.* **2022**, *22*, 285–315. https://doi.org/10.1007/s13127-021-00536-6.
107. Hancock, Z.B.; Hardin, F.O.; Murthy, A.; Hillhouse, A.; Johnston, J.S. Rapid Genomic Expansion and Purging Associated with Habitat Transitions in a Clade of Beach Crustaceans (Amphipoda: Haustoriidae). *J. Crustac. Biol.* **2021**, *41*, ruab042. https://doi.org/10.1093/jcbiol/ruab042.
108. Flot, J.-F.; Wörheide, G.; Dattagupta, S. Unsuspected Diversity of *Niphargus* Amphipods in the Chemoautotrophic Cave Ecosystem of Frasassi, Central Italy. *BMC Evol. Biol.* **2010**, *10*, 171. https://doi.org/10.1186/1471-2148-10-171.
109. Saranchina, A.; Mutin, A.; Govorukhina, E.; Rzhechitskiy, Y.; Gurkov, A.; Timofeyev, M.; Drozdova, P. Genetic Diversity in a Baikal Species Complex *Eulimnogammarus verrucosus* (Amphipoda: Gammaroidea) in the Angara River, the Only Outflow of Lake Baikal. *Zool. Scr.* **2024**, *53*, 867–879. https://doi.org/10.1111/zsc.12681.
110. Gurkov, A.; Rivarola-Duarte, L.; Bedulina, D.; Fernández Casas, I.; Michael, H.; Drozdova, P.; Nazarova, A.; Govorukhina, E.; Timofeyev, M.; Stadler, P.F.; et al. Indication of Ongoing Amphipod Speciation in Lake Baikal by Genetic Structures Within Endemic Species. *BMC Evol. Biol.* **2019**, *19*, 138. https://doi.org/10.1186/s12862-019-1470-8.
111. Drozdova, P.; Saranchina, A.; Madyarova, E.; Gurkov, A.; Timofeyev, M. Experimental Crossing Confirms Reproductive Isolation between Cryptic Species Within *Eulimnogammarus verrucosus* (Crustacea: Amphipoda) from Lake Baikal. *Int. J. Mol. Sci.* **2022**, *23*, 10858. https://doi.org/10.3390/ijms231810858.
112. Jakob, L.; Vereshchagina, K.P.; Tillmann, A.; Rivarola-Duarte, L.; Axenov-Gribanov, D.V.; Bedulina, D.S.; Gurkov, A.N.; Drozdova, P.; Timofeyev, M.A.; Stadler, P.F.; et al. Thermal Reaction Norms of Key Metabolic Enzymes Reflect Divergent Physiological and Behavioral Adaptations of Closely Related Amphipod Species. *Sci. Rep.* **2021**, *11*, 4562. https://doi.org/10.1038/s41598-021-83748-2.
113. Lu, D.; Liu, H.; Tong, Y.; Liu, Z.; Zhu, C.-D.; Hou, Z. An Annotated Genome of Freshwater Amphipod (*Gammarus nekkensis*). *Sci. Data* **2026**, *13*, 790. https://doi.org/10.1038/s41597-026-07126-1.
114. Poynton, H.C.; Hasenbein, S.; Benoit, J.B.; Sepulveda, M.S.; Poelchau, M.F.; Hughes, D.S.T.; Murali, S.C.; Chen, S.; Glastad, K.M.; Goodisman, M.A.D.; et al. The Toxicogenome of *Hyalella azteca*: A Model for Sediment Ecotoxicology and Evolutionary Toxicology. *Environ. Sci. Technol.* **2018**, *52*, 6009–6022. https://doi.org/10.1021/acs.est.8b00837.
115. Kao, D.; Lai, A.G.; Stamataki, E.; Rosic, S.; Konstantinides, N.; Jarvis, E.; Di Donfrancesco, A.; Pouchkina-Stancheva, N.; Sémon, M.; Grillo, M.; et al. The Genome of the Crustacean *Parhyale hawaiensis*, a Model for Animal Development, Regeneration, Immunity and Lignocellulose Digestion. *eLife* **2016**, *5*, e20062. https://doi.org/10.7554/eLife.20062.
116. Rand, D.M.; Nunez, J.C.B.; Williams, S.; Rong, S.; Burley, J.T.; Neil, K.B.; Spierer, A.N.; McKerrow, W.; Johnson, D.S.; Raynes, Y.; et al. Parasite Manipulation of Host Phenotypes Inferred from Transcriptional Analyses in a Trematode-Amphipod System. *Mol. Ecol.* **2023**, *32*, 5028–5041. https://doi.org/10.1111/mec.17093.
117. Patra, A.K.; Chung, O.; Yoo, J.Y.; Kim, M.S.; Yoon, M.G.; Choi, J.-H.; Yang, Y. First Draft Genome for the Sand-Hopper *Trinorchestia longiramus*. *Sci. Data* **2020**, *7*, 85. https://doi.org/10.1038/s41597-020-0424-8.
118. Telnes, E.; Mutin, A.; Rzhechitskiy, Y.; Shirokova, Y.; Zolotovskaya, E.; Kondrateva, E.; Govorukhina, E.; Gurkov, A.; Timofeyev, M.; Drozdova, P. The Gems of Lake Baikal: Scavenger Amphipods of the Genus *Ommatogammarus* Sensu Lato Revisited. *Zoologica Scripta* **2026**, https://doi.org/10.1111/zsc.70102.
119. Bakker, F.T.; Lei, D.; Yu, J.; Mohammadin, S.; Wei, Z.; van de Kerke, S.; Gravendeel, B.; Nieuwenhuis, M.; Staats, M.; Alquezar-Planas, D.E.; et al. Herbarium Genomics: Plastome Sequence Assembly from a Range of Herbarium Specimens Using an Iterative Organelle Genome Assembly Pipeline. *Biol. J. Linn. Soc.* **2016**, *117*, 33–43. https://doi.org/10.1111/bij.12642.
120. Plese, B.; Rossi, M.E.; Kenny, N.J.; Taboada, S.; Koutsouveli, V.; Riesgo, A. Trimitomics: An Efficient Pipeline for Mitochondrial Assembly from Transcriptomic Reads in Nonmodel Species. *Mol. Ecol. Resour.* **2019**, *19*, 1230–1239. https://doi.org/10.1111/1755-0998.13033.

121. The Galaxy Community Galaxy for Accessible, Reproducible, and Collaborative Data Analyses: 2026 Update. *Nucleic Acids Res.* **2026**, *54*, W105–W116. https://doi.org/10.1093/nar/gkag469.
122. Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. Using SPAdes de Novo Assembler. *Curr. Protoc. Bioinform.* **2020**, *70*, e102. https://doi.org/10.1002/cpbi.102.
123. Wheeler, T.J.; Eddy, S.R. Nhmmer: DNA Homology Search with Profile HMMs. *Bioinformatics* **2013**, *29*, 2487–2489. https://doi.org/10.1093/bioinformatics/btt403.
124. Warren, R.L.; Sutton, G.G.; Jones, S.J.M.; Holt, R.A. Assembling Millions of Short DNA Sequences Using SSAKE. *Bioinformatics* **2007**, *23*, 500–501. https://doi.org/10.1093/bioinformatics/btl629.
125. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; 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.
126. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A.D. rnaSPAdes: A de Novo Transcriptome Assembler and Its Application to RNA-Seq Data. *Gigascience* **2019**, *8*, giz100. https://doi.org/10.1093/gigascience/giz100.
127. Chevreux, B.; Wetter, T.; Suhai, S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. In *Proceedings of the German Conference on Bioinformatics*; Citeseer: University Park, PA, USA, 1999; Volume 99, pp. 45–56.
128. Li, D.; Luo, R.; Liu, C.-M.; Leung, C.-M.; Ting, H.-F.; Sadakane, K.; Yamashita, H.; Lam, T.-W. MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices. *Methods* **2016**, *102*, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
129. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res.* **2017**, *27*, 824–834. https://doi.org/10.1101/gr.213959.116.
130. Peng, Y.; Leung, H.C.M.; Yiu, S.M.; Chin, F.Y.L. IDBA—A Practical Iterative de Bruijn Graph De Novo Assembler. In *Proceedings of the Research in Computational Molecular Biology*; Berger, B., Ed.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 426–440.
131. Xie, Y.; Wu, G.; Tang, J.; Luo, R.; Patterson, J.; Liu, S.; Huang, W.; He, G.; Gu, S.; Li, S.; et al. SOAPdenovo-Trans: De Novo Transcriptome Assembly with Short RNA-Seq Reads. *Bioinformatics* **2014**, *30*, 1660–1666. https://doi.org/10.1093/bioinformatics/btu077.
132. Simpson, J.T.; Wong, K.; Jackman, S.D.; Schein, J.E.; Jones, S.J.M.; Birol, İ. ABySS: A Parallel Assembler for Short Read Sequence Data. *Genome Res.* **2009**, *19*, 1117–1123. https://doi.org/10.1101/gr.089532.108.
133. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
134. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. Available online: https://bbmap.org (accessed on 24 June 2026).
135. Li, H. Seqtk. Available online: https://github.com/lh3/seqtk (accessed on 30 July 2026).
136. Freudenthal, J.A.; Pfaff, S.; Terhoeven, N.; Korte, A.; Ankenbrand, M.J.; Förster, F. A Systematic Comparison of Chloroplast Genome Assembly Tools. *Genome Biol.* **2020**, *21*, 254. https://doi.org/10.1186/s13059-020-02153-6.
137. Li, H. New Strategies to Improve Minimap2 Alignment Accuracy. *Bioinformatics* **2021**, *37*, 4572–4574. https://doi.org/10.1093/bioinformatics/btab705.
138. Quinlan, A.R.; Hall, I.M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
139. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
140. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
141. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.
142. Heger, A.; Marshall, J.; Jacobs, K. Pysam: Htslib Interface for Python. Available online: https://pysam.readthedocs.io/en/latest/ (accessed on 9 October 2026).
143. Bonfield, J.K.; Marshall, J.; Danecek, P.; Li, H.; Ohan, V.; Whitwham, A.; Keane, T.; Davies, R.M. HTSlib: C Library for Reading/Writing High-Throughput Sequencing Data. *Gigascience* **2021**, *10*, giab007. https://doi.org/10.1093/gigascience/giab007.
144. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

145. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. https://doi.org/10.1109/MCSE.2007.55.
146. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinform.* **2009**, *10*, 421. https://doi.org/10.1186/1471-2105-10-421.
147. Jung, J.; Kim, J.I.; Jeong, Y.-S.; Yi, G. AGORA: Organellar Genome Annotation from the Amino Acid and Nucleotide References. *Bioinformatics* **2018**, *34*, 2661–2663. https://doi.org/10.1093/bioinformatics/bty196.
148. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and Accurate Annotation of Organelle Genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. https://doi.org/10.1093/nar/gkx391.
149. Donath, A.; Jühling, F.; Al-Arab, M.; Bernhart, S.H.; Reinhardt, F.; Stadler, P.F.; Middendorf, M.; Bernt, M. Improved Annotation of Protein-Coding Genes Boundaries in Metazoan Mitochondrial Genomes. *Nucleic Acids Res.* **2019**, *47*, 10543–10552. https://doi.org/10.1093/nar/gkz833.
150. Lang, B.F.; Beck, N.; Prince, S.; Sarrasin, M.; Rioux, P.; Burger, G. Mitochondrial Genome Annotation with MFannot: A Critical Analysis of Gene Identification and Gene Model Prediction. *Front. Plant Sci.* **2023**, *14*, 1222186. https://doi.org/10.3389/fpls.2023.1222186.
151. Kent, W.J. BLAT—The BLAST-like Alignment Tool. *Genome Res.* **2002**, *12*, 656–664. https://doi.org/10.1101/gr.229202.
152. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. https://doi.org/10.1093/nar/gkr367.
153. Durbin, R.; Eddy, S.; Krogh, A.; Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, UK, 1998; ISBN 0-521-62041-4.
154. Jühling, F.; Pütz, J.; Bernt, M.; Donath, A.; Middendorf, M.; Florentz, C.; Stadler, P.F. Improved Systematic tRNA Gene Annotation Allows New Insights into the Evolution of Mitochondrial tRNA Structures and into the Mechanisms of Mitochondrial Genome Rearrangements. *Nucleic Acids Res.* **2012**, *40*, 2833–2845. https://doi.org/10.1093/nar/gkr1131.
155. Chan, P.P.; Lin, B.Y.; Mak, A.J.; Lowe, T.M. tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *Nucleic Acids Res.* **2021**, *49*, 9077–9096. https://doi.org/10.1093/nar/gkab688.
156. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-Fold Faster RNA Homology Searches. *Bioinformatics* **2013**, *29*, 2933–2935. https://doi.org/10.1093/bioinformatics/btt509.
157. Laslett, D.; Canbäck, B. ARWEN: A Program to Detect tRNA Genes in Metazoan Mitochondrial Nucleotide Sequences. *Bioinformatics* **2008**, *24*, 172–175. https://doi.org/10.1093/bioinformatics/btm573.
158. Laslett, D.; Canback, B. ARAGORN, a Program to Detect tRNA Genes and tmRNA Genes in Nucleotide Sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. https://doi.org/10.1093/nar/gkh152.
159. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic Annotation of Organellar Genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. https://doi.org/10.1093/bioinformatics/bth352.
160. Sheffield, N.C.; Hiatt, K.D.; Valentine, M.C.; Song, H.; Whiting, M.F. Mitochondrial Genomics in Orthoptera Using MOSAS. *Mitochondrial DNA* **2010**, *21*, 87–104. https://doi.org/10.3109/19401736.2010.500812.
161. Fiedler, L.; Middendorf, M.; Bernt, M. Fully Automated Annotation of Mitochondrial Genomes Using a Cluster-Based Approach with de Bruijn Graphs. *Front. Genet.* **2023**, *14*, 1250907. https://doi.org/10.3389/fgene.2023.1250907.
162. Fiedler, L.; Bernt, M.; Middendorf, M. DeGeCI 1.1: A Web Platform for Gene Annotation of Mitochondrial Genomes. *Bioinform. Adv.* **2024**, *4*, vbae072. https://doi.org/10.1093/bioadv/vbae072.
163. Carbajosa, G.; Ali, A.T.; Hodgkinson, A. Identification of Human Mitochondrial RNA Cleavage Sites and Candidate RNA Processing Factors. *BMC Biol.* **2022**, *20*, 168. https://doi.org/10.1186/s12915-022-01373-5.
164. Schmieder, R.; Edwards, R. Quality Control and Preprocessing of Metagenomic Datasets. *Bioinformatics* **2011**, *27*, 863–864. https://doi.org/10.1093/bioinformatics/btr026.
165. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
166. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Use R!; Springer: Cham, Switzerland, 2016; ISBN 978-3-319-24277-4.
167. Wilkins, D. Gggenes: Draw Gene Arrow Maps in "Ggplot2". Available online: https://wilkox.org/gggenes/ (accessed on 9 October 2026).
168. Wickham, H.; Hester, J.; Bryan, J. Readr: Read Rectangular Text Data. Available online: https://readr.tidyverse.org (accessed on 1 October 2026).
169. Tan, E.Y.W.; Quek, Z.B.R.; Neo, M.L.; Fauvelot, C.; Huang, D. Genome Skimming Resolves the Giant Clam (Bivalvia: Cardiidae: Tridacninae) Tree of Life. *Coral Reefs* **2022**, *41*, 497–510. https://doi.org/10.1007/s00338-020-02039-w.

170. Uribe, J.E.; Nava, S.; Murphy, K.R.; Tarragona, E.L.; Castro, L.R. Characterization of the Complete Mitochondrial Genome of *Amblyomma ovale*, Comparative Analyses and Phylogenetic Considerations. *Exp. Appl. Acarol.* **2020**, *81*, 421–439. https://doi.org/10.1007/s10493-020-00512-3.
171. Song, N.; An, S.; Yin, X.; Cai, W.; Li, H. Application of RNA-Seq for Mitogenome Reconstruction, and Reconsideration of Long-Branch Artifacts in Hemiptera Phylogeny. *Sci. Rep.* **2016**, *6*, 33465. https://doi.org/10.1038/srep33465.
172. de Siqueira e Silva, M.C.; Guadanucci, J.P.L.; Belintani, T. Mitochondrial Genomes from RNA-Seq Reveal Phylogeny and Selection in *Mepraia* (Hemiptera: Reduviidae). *Mol. Genet. Genom.* **2026**, *301*, 117. https://doi.org/10.1007/s00438-026-02434-y.
173. Torson, A.S.; Hicks, A.M.A.; Baragar, C.E.; Smith, D.R.; Sinclair, B.J. The Mitochondrial Genomes of Two *Gryllus* Crickets (Grylloidea: Gryllidae) via RNA-Seq. *Mitochondrial DNA Part B* **2022**, *7*, 106–107. https://doi.org/10.1080/23802359.2021.2010613.
174. Bae, S. Relative Performance of Three Phylogenetic Methods Based on Complete Mitochondrial Genomes of Barnacle. *Sci. Rep.* **2025**, *15*, 34799. https://doi.org/10.1038/s41598-025-18483-z.
175. Sirotinina, E.A.; Sherbakov, D.Y.; Romanova, E.V. Evolution of Gene Order in mtDNA of Baikal Endemic Amphipods and Its Possible Mechanisms. *Vavilovskii Zhurnal Genet. Sel.* **2026**, *30*, 136–145. https://doi.org/10.18699/vjgb-26-15.
176. Goodman, S.J.; Carr, I.M. Drawing Mitochondrial Genomes with circularMT. *Bioinformatics* **2024**, *40*, btae450. https://doi.org/10.1093/bioinformatics/btae450.
177. Milián-García, Y.; Hempel, C.A.; Janke, L.A.A.; Young, R.G.; Furukawa-Stoffer, T.; Ambagala, A.; Steinke, D.; Hanner, R.H. Mitochondrial Genome Sequencing, Mapping, and Assembly Benchmarking for Culicoides Species (Diptera: Ceratopogonidae). *BMC Genom.* **2022**, *23*, 584. https://doi.org/10.1186/s12864-022-08743-x.
178. Calabrese, C.; Simone, D.; Diroma, M.A.; Santorsola, M.; Guttà, C.; Gasparre, G.; Picardi, E.; Pesole, G.; Attimonelli, M. MToolBox: A Highly Automated Pipeline for Heteroplasmy Annotation and Prioritization Analysis of Human Mitochondrial Variants in High-Throughput Sequencing. *Bioinformatics* **2014**, *30*, 3115–3117. https://doi.org/10.1093/bioinformatics/btu483.
179. Mahar, N.S.; Satyam, R.; Sundar, D.; Gupta, I. A Systematic Comparison of Human Mitochondrial Genome Assembly Tools. *BMC Bioinform.* **2023**, *24*, 341. https://doi.org/10.1186/s12859-023-05445-3.
180. Ni, Y.; Li, J.; Tan, Y.; Shen, G.; Liu, C. Advance in the Assembly of the Plant Mitochondrial Genomes Using High-Throughput DNA Sequencing Data of Total Cellular DNAs. *Plant Biotechnol. J.* **2025**, *23*, 4944–4965. https://doi.org/10.1111/pbi.70249.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.