# Variants of the Mitochondrial Genome Across Species

A comparative survey of protein-coding, rRNA and tRNA gene complements, control-region architecture, and every other known axis of mitogenomic variation

Reference compilation

Prepared 19 August 2026

*Figures are drawn from the published comparative-mitogenomics literature and are approximate: assembly method, strain and annotation pipeline all shift reported values. See §11 for verification sources.*

# Contents

*If the contents list below appears blank, select it and press F9 (Word) to populate page numbers.*

# 1. Scope, method and how to read this document

The phrase "the mitochondrial genome" is a vertebrate parochialism. The 16.5 kb circle with 37 genes that appears in every textbook is one of the most derived, most compacted and least representative mitochondrial genomes in the eukaryotic domain. Across the full breadth of eukaryotes, mitochondrial DNA (mtDNA) varies by more than three orders of magnitude in size, from roughly 6 kb to almost 19 Mb; by two orders of magnitude in gene number, from 2 to 100; in topology from a single circle to a linear chromosome with telomeres, to thousands of catenated rings, to eighty separate mini-chromosomes; and in genetic code across at least thirteen distinct translation tables. In several lineages it has been lost outright.

This document catalogues that variation systematically. It is organised in two complementary passes:

- **Axis-by-axis (§4).** Each structural or informational feature of a mitochondrial genome — topology, size, protein complement, rRNA genes, tRNA genes, the control region, gene order, genetic code, introns, RNA editing, repeats, copy number, nuclear transfer, cryptic ORFs — is treated in turn, with the full observed range of states and the taxa that occupy them.
- **Taxon-by-taxon (§5).** The same information re-cut by lineage, so that the mitogenome of any given group can be looked up as a coherent whole.

Sections 2 and 3 establish the reference architecture and the analytical framework. Section 6 tabulates the extremes. Section 7 draws out the points of contact with ageing biology, where cross-species mitogenomic variation is a persistently under-exploited natural experiment. Sections 8–11 cover pitfalls, nomenclature and primary sources.

## 1.1 A note on "exhaustive"

An exhaustive list of mtDNA variants at the level of individual species is not achievable in a document of any length: NCBI's organelle collection holds tens of thousands of complete mitogenomes, and the number grows weekly. What is achievable, and what this document attempts, is exhaustiveness at the level of variant types — every known structural, genic and informational way in which a mitochondrial genome can differ from the vertebrate baseline, with named exemplar taxa for each. Where a state is known from only one or two organisms, they are named; where it is a general property of a clade, the clade is named and representative sizes are given.

Two caveats worth carrying throughout. First, gene presence in a mitogenome is not the same as gene presence in the organism: almost every gene "lost" from mtDNA has been relocated to the nucleus rather than abolished, and the protein is imported. Second, "circular" is very often a property of the assembly rather than the molecule; in vivo, most mtDNA that maps as a circle exists as linear concatemers or branched recombination intermediates.

# 2. The reference architecture: the vertebrate mitochondrial genome

Human mtDNA (the revised Cambridge Reference Sequence, rCRS, NC_012920) is 16,569 bp of double-stranded circular DNA carrying 37 genes: 13 protein-coding, 22 tRNA and 2 rRNA. It is extraordinarily compact. There are essentially no intergenic spacers, no introns, and in several places genes overlap: ATP8 and ATP6 share 46 nucleotides in different frames, and ND4L overlaps ND4. Two genes (ND6 and eight tRNAs) sit on the light strand; the remaining 28 are on the heavy strand.

The two strands are distinguished by buoyant density in caesium chloride, a consequence of asymmetric base composition: the heavy (H) strand is guanine-rich, the light (L) strand cytosine-rich. This asymmetry is itself a product of the replication mechanism, which leaves the H-strand template single-stranded and therefore exposed to spontaneous cytosine deamination for extended periods.

Transcription is polycistronic. Both strands are transcribed almost in their entirety from promoters in the control region (LSP and HSP), and the primary transcripts are then cut into individual mRNAs and rRNAs by excision of the interspersed tRNAs — the "tRNA punctuation" model of Ojala, Montoya and Attardi (1981). This is why the 22 tRNA genes are distributed around the circle rather than clustered: they function as processing signals as much as adaptors. Several mRNAs are not terminated by a complete stop codon in the DNA at all, but by an incomplete T or TA that is converted to UAA by post-transcriptional polyadenylation.

## 2.1 The 37 genes of human mtDNA

| Gene | Class | Product / role | Strand | Approx. size |
|---|---|---|---|---|
| MT-ND1 | Protein | Complex I subunit ND1 | H | 956 bp |
| MT-ND2 | Protein | Complex I subunit ND2 | H | 1,042 bp |
| MT-ND3 | Protein | Complex I subunit ND3 | H | 346 bp |
| MT-ND4L | Protein | Complex I subunit ND4L (overlaps ND4) | H | 297 bp |
| MT-ND4 | Protein | Complex I subunit ND4 | H | 1,378 bp |
| MT-ND5 | Protein | Complex I subunit ND5 | H | 1,812 bp |
| MT-ND6 | Protein | Complex I subunit ND6 — the sole L-strand protein gene | L | 525 bp |
| MT-CYB | Protein | Complex III cytochrome b | H | 1,141 bp |
| MT-CO1 | Protein | Complex IV subunit I (catalytic; DNA-barcode locus) | H | 1,542 bp |
| MT-CO2 | Protein | Complex IV subunit II (CuA centre) | H | 684 bp |
| MT-CO3 | Protein | Complex IV subunit III | H | 784 bp |
| MT-ATP6 | Protein | Complex V Fo subunit a | H | 681 bp |
| MT-ATP8 | Protein | Complex V Fo subunit A6L (overlaps ATP6) | H | 207 bp |
| MT-RNR1 | rRNA | 12S small-subunit rRNA (mt-SSU) | H | 954 bp |
| MT-RNR2 | rRNA | 16S large-subunit rRNA (mt-LSU); hosts humanin ORF | H | 1,559 bp |
| 22 MT-T* genes | tRNA | One tRNA per amino acid, plus two each for Leu and Ser | 14 H / 8 L | 59–75 bp each |
| D-loop / CR | Non-coding | Origin of H-strand replication, both promoters, 7S DNA | — | 1,122 bp |

*Table 1. Gene complement of the canonical vertebrate mitochondrial genome. Complex II (succinate dehydrogenase) is entirely nucleus-encoded in animals — a lineage-specific loss, since sdh subunits are retained in mtDNA of plants and many protists.*

## 2.2 The D-loop and control region in detail

The D-loop is the only substantial non-coding tract in vertebrate mtDNA and the only region that is regularly reproduced as a three-stranded structure. Its name comes from the displacement loop formed when a short nascent H-strand — 7S DNA, roughly 650 nucleotides — is synthesised from the origin of H-

strand replication (OH) and remains annealed to the L-strand template, displacing the parental H-strand. In most cells the majority of mtDNA molecules carry this triplex at any given moment.

Functionally it is not one element but a cluster of them packed into about 1.1 kb:

- **OH (origin of H-strand replication)** — nucleotide 191 in human numbering, where the transition from RNA primer to DNA occurs.
- **HSP1 and HSP2 (heavy-strand promoters)** — HSP1 drives a short transcript covering the two rRNAs and two tRNAs; HSP2 drives the near-genome-length polycistron.
- **LSP (light-strand promoter)** — bidirectional partner to HSP; its transcript also supplies the RNA primer for OH after processing by RNase MRP at the CSBs.
- **CSB1, CSB2, CSB3 (conserved sequence blocks)** — primer-processing and G-quadruplex-forming elements. CSB2 forms a stable G-quadruplex R-loop that determines where transcription switches to replication priming.
- **TAS / ETAS (termination-associated sequences)** — where 7S DNA synthesis terminates, defining the 3' end of the D-loop.
- **mt5 and other protein-binding sites** — TFAM, MTERF1 and mtSSB footprints; TFAM binding here is both packaging and transcriptional activation.
- **HVR1, HVR2, HVR3 (hypervariable regions)** — the fastest-evolving segments of the animal mitogenome and the substrate of nearly all forensic and phylogeographic mtDNA typing. HVR1 spans roughly np 16024–16365, HVR2 np 57–372.

Three further features matter for anyone using D-loop data. It contains variable-number tandem repeats in many taxa, which cause length heteroplasmy and can confound alignment. Its mutation rate is roughly 5–10 times the coding-region average, so it saturates quickly and is unsuitable for deep phylogeny. And in some lineages the control region has been duplicated, translocated, or split — discussed in §4.6.

## 2.3 How far the vertebrate pattern extends

The 37-gene, 13-protein set is stable across essentially all vertebrates and most bilaterian animals, which is why it is easy to mistake for universal. Within vertebrates, variation is confined to control-region duplications, occasional tRNA rearrangements, and the ND6/tRNA cluster inversions used as phylogenetic markers. Outside Bilateria, and entirely outside Metazoa, almost every feature listed above becomes negotiable.

## 3. The axes of variation: an analytical framework

Mitogenomic diversity is easier to hold in mind if it is decomposed into independent axes. A given lineage occupies a position on each; the axes are only loosely correlated, which is why "reduced" and "expanded" mitogenomes are not simply opposite ends of a single scale. A dinoflagellate mitogenome is simultaneously gene-poor and physically enormous; a jakobid one is gene-rich and physically small.

| # | Axis | Range of observed states |
|---|---|---|
| 1 | Topology and physical form | Circular monomer; linear with telomeres; linear concatemer; branched network; multipartite (2 to thousands of chromosomes); catenated kinetoplast; absent |
| 2 | Genome size | ~5.9 kb (Plasmodium) to 18.99 Mb (Cathaya) — a 3,200-fold range |
| 3 | Protein-coding complement | 0 to 66 genes; complex I present, partially lost, or wholly lost; ribosomal-protein and cytochrome-maturation genes present or nuclear |
| 4 | rRNA genes | Two intact; fragmented into modules; scrambled and dispersed; reduced to <500 nt; 5S present or absent or functionally replaced |
| 5 | tRNA genes | 0 to 34; canonical cloverleaf, D-arm-less, T-arm-less, or fully armless; wholesale cytosolic import in several lineages |
| 6 | Control region | Single D-loop; duplicated; split; tandem-repeat-laden; multiple dispersed promoters; no discrete control region at all |
| 7 | Gene order and synteny | Deeply conserved (vertebrates) to hypervariable (molluscs, hymenopterans, plants) |
| 8 | Genetic code | Standard plus at least twelve mitochondrial reassignments (NCBI tables 2–5, 9, 13, 14, 16, 21–24, 33) |
| 9 | Introns and mobile elements | None (vertebrates); group I; group II; both; intron-encoded maturases and homing endonucleases; plasmid-derived and viral insertions |
| 10 | RNA editing and processing | None; C-to-U and U-to-C substitution; U-insertion/deletion; oligo-U 3' addition; nucleotide insertion; tRNA 5' editing; cis- and trans-splicing |
| 11 | Repeats and non-coding DNA | <1% (animals) to >95% (Pinaceae, Silene); direct and inverted repeats driving recombination |
| 12 | Copy number, heteroplasmy, inheritance | 100 to 100,000+ copies per cell; homoplasmy to persistent heteroplasmy; maternal, paternal, biparental, or doubly uniparental |
| 13 | Nuclear and organellar transfer | NUMTs; mitochondrion-to-plastid and plastid-to-mitochondrion transfer; horizontal acquisition of entire foreign mitogenomes |
| 14 | Cryptic ORFs and derived peptides | Humanin, MOTS-c, SHLP1–6, sORFs; unassigned ymf/orf genes in fungi and protists |

*Table 2. The fourteen axes along which mitochondrial genomes vary. Each is treated in turn in §4.*

# 4. Variation axis by axis

## 4.1 Topology and physical form

The textbook circle is common but far from universal, and even where a genome assembles as a circle the in vivo molecule is frequently something else. Southern blotting and pulsed-field gel work has repeatedly shown that "circular" mtDNA in fungi, plants and even animals exists largely as head-to-tail linear concatemers and branched recombination intermediates, with true monomeric circles a minority species.

| Form | Description | Exemplar taxa |
|---|---|---|
| Circular monomer | The canonical form; single covalently closed circle | Vertebrates, most arthropods, most angiosperms (as a mapping construct), Anthozoa |
| Linear with terminal telomeres | Defined ends, often with inverted-repeat telomeres and covalently closed hairpins | Medusozoan cnidarians (Hydra, Aurelia), ciliates (Tetrahymena, Paramecium), Chlamydomonas, several yeasts (Candida, Pichia), some Chytridiomycota |
| Linear concatemer / branched network | Multigenomic head-to-tail arrays and recombination-generated branches; the dominant in vivo state in plants and many fungi | Land plants, Saccharomyces, Neurospora |
| Multipartite — few chromosomes | Genome split across a small number of circles or linear elements | Cucumis melo (3 circles), Silene spp., Amborella, rotifers (Brachionus, 2 chromosomes), Globodera nematodes |
| Multipartite — many minicircles | Genome fragmented into dozens of small circles, each carrying one or few genes | Sucking lice (Pediculus humanus, ~18 minicircles of 3–4 kb); some ticks; diplonemids (~80 circular chromosomes) |
| Kinetoplast (catenated network) | A few dozen maxicircles plus 5,000–10,000 minicircles interlocked into a single disc-shaped network | Trypanosoma, Leishmania, Crithidia and other kinetoplastids |
| Tandem linear repeats | Short unit repeated head-to-tail into long linear arrays | Plasmodium and other Apicomplexa (~6 kb unit) |
| Mixed linear/circular isoforms | Linear monomer coexisting with circular dimer in the same individual | Armadillidium vulgare (terrestrial isopod) |
| No genome; organelle retained | Mitochondrion-derived organelle without DNA | Hydrogenosomes (Trichomonas), mitosomes (Entamoeba, Giardia, Microsporidia), Henneguya salminicola (a cnidarian — the first animal shown to lack mtDNA) |
| No organelle at all | Complete secondary loss of the mitochondrion | Monocercomonoides sp. PA203 |

*Table 3. Physical forms of mitochondrial DNA.*

## 4.2 Genome size

Size and gene content are decoupled almost completely. The 18.99 Mb Cathaya argyrophylla mitogenome carries roughly the same functional gene set as the 66 kb mitogenome of the mistletoe Viscum scurruloideum, and vastly less than the 69 kb genome of Reclinomonas americana. Expansion in plants is driven by repeats, transposable-element activity, and large-scale import of plastid and nuclear sequence, not by gene acquisition.

| Organism | Size | Remark |
|---|---|---|
| Plasmodium falciparum | ~5.97 kb | Smallest known mitogenome; 3 protein genes, fragmented rRNA, no tRNAs |
| Mnemiopsis leidyi (ctenophore) | ~10 kb | Smallest animal mitogenome; no tRNAs, drastically truncated rRNAs |
| Chlamydomonas reinhardtii | ~15.8 kb | Linear; 8 genes; scrambled rRNA modules |
| Homo sapiens | 16,569 bp | The reference |
| Caenorhabditis elegans | 13,794 bp | atp8 absent |
| Drosophila melanogaster | ~19.5 kb | Large AT-rich control region |
| Schizosaccharomyces pombe | ~19.4 kb | |
| Trypanosoma brucei maxicircle | ~22 kb | Plus ~10,000 minicircles of ~1 kb each |
| Trichoplax adhaerens (Placozoa) | ~43 kb | Largest animal mitogenome; introns and extra ORFs |
| Viscum scurruloideum | ~66 kb | Smallest angiosperm mitogenome; all nad genes lost |
| Reclinomonas americana | ~69 kb | 97 genes — bacterial-like operons and RNA polymerase |
| Neurospora crassa | ~64.8 kb | Group I intron-rich |
| Saccharomyces cerevisiae | ~85.8 kb | ~80% AT; group I and II introns; complex I entirely absent |
| Arabidopsis thaliana | ~367 kb | Typical small angiosperm |
| Brassica napus | ~222 kb | Among the most compact angiosperm mitogenomes |
| Zea mays | ~570 kb | Multiple recombining subgenomic circles |
| Cucumis melo (melon) | ~2.9 Mb | Three chromosomes; short-repeat expansion |
| Amborella trichopoda | ~3.9 Mb | Six genome-equivalents of horizontally acquired foreign mtDNA |
| Silene noctiflora | ~6.7 Mb | Multichromosomal, ~60 circles |
| Curcuma longa | ~7.7 Mb | Network-like 12-contig structure |
| Silene conica | ~11.3 Mb | Mostly sequence of unknown origin |
| Larix sibirica | ~11.7 Mb | Previous record-holder |
| Cathaya argyrophylla | 18.99 Mb | Current record; extensive plastid-derived sequence and LINE/LTR activity |

*Table 4. Mitochondrial genome sizes across the eukaryotic range. Values are approximate and assembly-dependent; plant figures in particular vary with how recombination isoforms are counted.*

## 4.3 Protein-coding gene complement

The ancestral mitochondrial genome, reconstructed from jakobids, carried something close to 66 protein genes: the full oxidative-phosphorylation set, a large block of ribosomal proteins, cytochrome-c maturation machinery, a bacterial-type multisubunit RNA polymerase, and protein-export components. Everything since has been attrition, with the surviving core converging on the most hydrophobic, hardest-to-import subunits — consistent with the co-location for redox regulation (CoRR) hypothesis and with the "hydrophobicity barrier" argument for why relocation stalls.

| Gene / family | Vertebrates | Other animals | Fungi | Land plants | Jakobids | Apicomplexa |
|---|---|---|---|---|---|---|
| nad1–nad6, nad4L (Complex I) | All 7 | All 7 | Lost in S. cerevisiae; present in Neurospora, Yarrowia | All 7 (lost in Viscum) | All 7 | None |
| nad7, nad9, nad11 | Nuclear | Nuclear | Variable | nad7, nad9 present | Present | None |
| cob (Complex III) | Yes | Yes | Yes | Yes | Yes | Yes |
| cox1, cox2, cox3 (Complex IV) | All 3 | All 3 | All 3 | All 3 (cox2 nuclear in some legumes) | All 3 + cox11 | cox1, cox3 only |
| atp6 | Yes | Yes | Yes | Yes | Yes | No |
| atp8 | Yes | Lost in nematodes, some flatworms | Yes | Yes (atp8/orfB) | Yes | No |
| atp9 | Nuclear | Retained in Porifera, Placozoa | Yes | Yes | Yes | No |
| atp1, atp4 | Nuclear | Nuclear | Nuclear | atp1, atp4 present | Present | No |
| sdh2, sdh3, sdh4 (Complex II) | Nuclear | Nuclear | Nuclear | sdh3, sdh4 variable | All present | No |
| ccmB, ccmC, ccmF (cyt c maturation) | Nuclear | Nuclear | Nuclear | Present | Present | No |
| rps/rpl (ribosomal proteins) | None | None | var1 (S. cerevisiae) only | ~15–20, lineage-variable | ~27 | None |
| rpoA, rpoB, rpoC, rpoD | None | None | None | None | Present — unique to jakobids | None |
| tatC / mttB (protein export) | None | None | None | Present | Present | None |
| secY, tufA | None | None | None | None | Present | None |
| matR (maturase) | None | None | Intron-encoded | Present | Absent | None |
| polB (DNA polymerase) | None | None | Linear-plasmid-derived in some | None | None | None |

*Table 5. Protein-coding gene presence and absence across major lineages. "Nuclear" indicates the gene exists but has been relocated to the nuclear genome and the product is imported.*

Notable individual cases. Medusozoan cnidarians carry polB, a DNA-polymerase gene of linear-plasmid ancestry, at the ends of their linear mitogenomes. Saccharomyces cerevisiae has lost the entire complex I gene set and compensates with alternative NADH dehydrogenases — the reason yeast is a poor model for complex I biology despite being an excellent one for mitochondrial genetics generally. The hemiparasitic mistletoe Viscum scurruloideum has likewise lost every nad gene and appears to lack a functional complex I entirely. Several legumes have transferred cox2 to the nucleus within the last few

million years, and in some species both copies are present simultaneously, capturing the transfer mid-event.

## 4.4 Ribosomal RNA genes

Two rRNA genes — small subunit and large subunit — are almost universal, but their integrity is not. Fragmentation is one of the most striking and least appreciated forms of mitogenomic variation.

- **Intact and contiguous.** Vertebrates (12S, 16S), most animals, fungi, land plants (18S, 26S, plus a 5S rRNA that animals lack).
- **Reduced.** Animal mt-rRNAs are already much shorter than their bacterial homologues, having lost most peripheral expansion segments. Ctenophores take this furthest: the Mnemiopsis mt-rRNAs are the shortest known.
- **Split into two pieces.** The LSU rRNA of many nematodes and of Tetrahymena is discontinuous.
- **Fragmented into many modules.** Chlamydomonas reinhardtii carries its two rRNAs as eight scrambled, interleaved modules dispersed around the linear genome, transcribed separately and assembled in trans. Dinoflagellates and Plasmodium do the same: the Plasmodium rRNAs exist as roughly twenty small fragments scattered across the 6 kb repeat unit in both orientations.
- **5S rRNA.** Present in land plants, jakobids, some algae and some protists; absent from animal and most fungal mitogenomes. In mammals the structural slot in the mitoribosome large subunit is occupied not by 5S rRNA but by mt-tRNA-Val (mt-tRNA-Phe in some species) — a striking example of an RNA repurposed as a structural scaffold, and a reason mt-tRNA-Val mutations produce ribosomal rather than translational phenotypes.

## 4.5 Transfer RNA genes

The tRNA complement is the single most variable gene class. The vertebrate set of 22 is the minimum sufficient for a self-contained mitochondrial translation system under a two-out-of-three wobble regime. Departures run in both directions.

| State | Count | Taxa and notes |
|---|---|---|
| Expanded | 24–34 | Land plants (some also import nuclear-encoded tRNAs); jakobids (Andalucia, 34 structural RNA genes); Saccharomyces (24) |
| Canonical animal set | 22 | Vertebrates and most bilaterians; includes duplicate tRNA-Leu (UUR/CUN) and tRNA-Ser (AGY/UCN) |
| Slightly reduced | 20–21 | Many nematodes, some molluscs — often with loss of one Ser or Leu isoacceptor |
| Severely reduced | 1–3 | Cnidaria: Anthozoa retain only trnM and trnW; Medusozoa often only trnM. Chlamydomonas retains 3. |
| Complete absence | 0 | Apicomplexa (Plasmodium), ctenophores (Mnemiopsis), kinetoplastid kDNA, dinoflagellates — all tRNAs imported from the cytosol |

*Table 6. Mitochondrial tRNA gene complements.*

### 4.5.1 Structurally aberrant tRNAs

Where tRNA genes are retained, their structure is often bizarre by cytosolic standards. Nematode mitochondrial tRNAs characteristically lack either the T-arm (replaced by a "TV-replacement loop") or the D-arm. In the enoplean nematode Romanomermis culicivorax and its relatives, tRNAs lacking both arms have been described — reduced almost to an acceptor stem and anticodon stem-loop, and functional only because a co-evolved, structurally divergent mitochondrial elongation factor Tu accommodates

them. Similar truncations occur in some arachnids, in the mite Varroa, and in molluscs. This is a clean demonstration that the tRNA cloverleaf is not a functional necessity but a historical constraint that can be relaxed when the translation apparatus co-evolves.

### 4.5.2 tRNA import

Wholesale import of nucleus-encoded tRNAs from the cytosol is the standard solution where mitochondrial tRNA genes have been lost. Kinetoplastids import their entire tRNA set. Apicomplexa likewise. Plants import a substantial subset (typically the tRNAs for Ala, Arg, Gly, Ile, Leu, Thr, Trp, Val), and the imported and native sets differ between species and even cultivars. Import routes vary: some use the protein-import channel of the TOM/TIM machinery, others a distinct carrier-mediated pathway. Human mitochondria have been reported to import small numbers of cytosolic tRNAs and other small RNAs (5S rRNA, RNase P RNA, MRP RNA), though the physiological significance remains contested.

### 4.5.3 tRNA editing

In Acanthamoeba castellanii and the chytrid Spizellomyces punctatus, mitochondrial tRNAs are transcribed with mismatched acceptor stems that are corrected by 5'-end editing — the first three nucleotides are replaced to restore pairing. A-to-I editing at the wobble position occurs in several lineages, and C-to-U editing of tRNAs is widespread in plants and marsupials, where it is required to generate a functional anticodon in tRNA-Asp.

## 4.6 The control region and its equivalents

The vertebrate D-loop is one solution to the problem of concentrating replication and transcription control in one place. It is not the only one, and in some lineages no discrete control region exists at all.

| Architecture | Taxa | Features |
|---|---|---|
| Single canonical D-loop | Most vertebrates | ~0.9–2 kb; OH, LSP, HSP, CSB1–3, TAS, HVR1–3; triple-stranded 7S DNA |
| Duplicated control region | Many birds (parrots, some Passeriformes), snakes (Boidae, Viperidae), some fishes, thrips, ticks | Two copies, usually kept near-identical by concerted evolution; sometimes one degenerates |
| Tandem-repeat-laden | Cats and other carnivores, sturgeon, salmonids, many insects | VNTRs produce length heteroplasmy within a single individual and complicate alignment |
| AT-rich region (insect equivalent) | Drosophila, Anopheles, Lepidoptera | Functionally analogous; extremely AT-biased; poly-T stretch and (TA)n repeats; can exceed 4 kb |
| Split or dispersed control elements | Some molluscs, some cnidarians | Replication origin and promoters separated at different genome positions |
| Multiple independent promoters, no single CR | Land plants | Dozens of scattered promoters; replication is recombination-dependent rather than origin-defined |
| ori/rep elements plus GC clusters | Saccharomyces cerevisiae | Seven or eight ori/rep sequences, of which only a few are active; GC clusters scattered through the AT-rich background |
| Telomeric origins | Medusozoa, ciliates, Chlamydomonas | Replication initiates from the linear termini; no internal D-loop |
| Minicircle conserved regions | Kinetoplastids | Each minicircle carries conserved sequence blocks; replication is network-wide and topologically coordinated |

| Architecture | Taxa | Features |
|---|---|---|
| No recognisable control region | Plasmodium, dinoflagellates | Rolling-circle or recombination-dependent replication of tandem repeats |

*Table 7. Control-region architectures.*

The practical consequence for anyone working with mtDNA sequence data is that "the control region" is not a comparable entity across taxa. Duplicated control regions in birds and snakes have repeatedly produced misleading phylogenies where the two copies were unknowingly aligned against each other, and length heteroplasmy in the repeat arrays of carnivores and salmonids is a routine source of sequencing artefacts.

## 4.7 Gene order and synteny

Gene order is one of the most useful phylogenetic characters available, precisely because rearrangement is rare enough to be informative but common enough to occur. Rates differ by orders of magnitude between lineages.

- **Near-frozen.** Vertebrate gene order is essentially invariant across ~500 Myr, with a small number of well-characterised exceptions: the tRNA-Ile/Gln/Met cluster rearrangement in birds, the "duplication-random loss" tRNA shuffles in reptiles, and control-region translocations in some fishes.
- **Moderately variable.** Most arthropods retain the ancestral pancrustacean order with sporadic tRNA translocations; these are strong clade markers.
- **Hypervariable.** Molluscs (particularly bivalves and gastropods), hymenopterans, nematodes and platyhelminths rearrange rapidly enough that gene order carries little signal above family level.
- **Effectively random.** Land plants recombine so freely across repeated sequences that gene order between congeners can differ completely while gene content is identical.
- **Operon-like conservation.** Jakobids preserve bacterial-style operons with Shine–Dalgarno-like ribosome-binding sites — the strongest surviving evidence of the alphaproteobacterial ancestry of the organelle.

Two mechanisms account for most animal rearrangement: tandem duplication followed by random loss of one copy of each duplicated gene (the TDRL model), and recombination between repeats. Inversions are less common in animals but dominate in plants.

## 4.8 Genetic code deviations

Mitochondria hold the great majority of known departures from the standard genetic code. The recurring themes are UGA reassigned from stop to tryptophan (which has happened independently many times), AUA reassigned from isoleucine to methionine, and various fates for the AGA/AGG codons.

| NCBI table | Name | Key reassignments | Taxa |
|---|---|---|---|
| 1 | Standard | None | Land plants, green algae (mostly), jakobids, Acanthamoeba, red algae |
| 2 | Vertebrate mitochondrial | UGA = Trp; AUA = Met; AGA, AGG = Stop | All vertebrates |
| 3 | Yeast mitochondrial | UGA = Trp; AUA = Met; CUN = Thr; CGA, CGC unassigned | Saccharomyces, Candida (some), other budding yeasts |

| NCBI table | Name | Key reassignments | Taxa |
|---|---|---|---|
| 4 | Mould, protozoan, coelenterate mitochondrial | UGA = Trp | Neurospora, Aspergillus, Anthozoa, Apicomplexa, Acanthamoeba, Mycoplasma (bacterial use) |
| 5 | Invertebrate mitochondrial | UGA = Trp; AUA = Met; AGA, AGG = Ser | Arthropods, molluscs, annelids, nematodes, most invertebrates |
| 9 | Echinoderm and flatworm mitochondrial | UGA = Trp; AGA, AGG = Ser; AAA = Asn | Echinoderms, most Platyhelminthes |
| 13 | Ascidian mitochondrial | UGA = Trp; AUA = Met; AGA, AGG = Gly | Tunicates (Ciona, Halocynthia) |
| 14 | Alternative flatworm mitochondrial | As table 9, plus UAA = Tyr | Some Platyhelminthes and nematodes |
| 16 | Chlorophycean mitochondrial | UAG = Leu | Some green algae (Chlorophyceae) |
| 21 | Trematode mitochondrial | UGA = Trp; AUA = Met; AGA, AGG = Ser; AAA = Asn | Trematode flatworms (Fasciola, Schistosoma) |
| 22 | Scenedesmus obliquus mitochondrial | UCA = Stop; UAG = Leu | Scenedesmus obliquus |
| 23 | Thraustochytrium mitochondrial | UUA = Stop | Thraustochytrium aureum |
| 24 | Rhabdopleuridae mitochondrial | AGA = Ser; AGG = Lys; UGA = Trp | Pterobranch hemichordates |
| 33 | Cephalodiscidae mitochondrial | AGA = Ser; AGG = Lys; UAA = Tyr; UGA = Trp | Cephalodiscus and relatives |

*Table 8. Mitochondrial genetic codes as catalogued by NCBI. Codes 6, 10, 12, 15 and others not listed apply to nuclear or bacterial systems.*

The mechanism of reassignment is generally either codon disappearance followed by reappearance with a new meaning (the "codon capture" model), or ambiguous intermediates in which a codon is read by two competing tRNAs until one is lost. The very high AT bias of many mitogenomes drives codons out of use and creates the conditions for capture, which is why reassignment is so much more frequent here than in nuclear genomes.

## 4.9 Introns and mobile elements

Vertebrate mtDNA is intronless. This is unusual. Group I and group II self-splicing introns are widespread elsewhere, and in some fungi they account for the majority of the genome.

- **Group I introns.** Fold into a conserved catalytic core requiring an exogenous guanosine cofactor. Abundant in fungi (Neurospora cox1 carries several; Podospora anserina has many), in Anthozoa (the cox1 and nad5 introns of corals and sea anemones), in green algae, and in some sponges and placozoans. Frequently host an ORF encoding a LAGLIDADG or GIY-YIG homing endonuclease that drives intron mobility by cleaving intron-less alleles — a genuinely selfish element that has been repurposed as the basis of much genome-editing technology.
- **Group II introns.** Splice via a lariat intermediate, mechanistically homologous to the nuclear spliceosome and almost certainly its ancestor. Dominant in land plants (roughly 20–26 in a typical angiosperm mitogenome) and present in fungi and some algae. Often encode a reverse-transcriptase/maturase (matR in plants).

- **Trans-splicing introns.** In plants, several group II introns are physically split across the genome, so that nad1, nad2 and nad5 are assembled from three to five separately transcribed pieces. Diplonemids take this to an extreme: essentially every gene is fragmented into modules on separate chromosomes and assembled in trans.
- **Plasmids and viral insertions.** Linear and circular mitochondrial plasmids (often carrying viral-type DNA and RNA polymerase genes) integrate into fungal and plant mitogenomes. The polB gene of medusozoan cnidarians and the "senDNA" senescence plasmids of Podospora anserina are examples with direct phenotypic consequences — the latter causing a heritable senescence syndrome that is one of the few clean examples of mitogenomic instability driving organismal ageing.

## 4.10 RNA editing and post-transcriptional processing

In several lineages the DNA sequence does not specify the protein sequence. Mitochondrial RNA editing is the most extreme deviation from colinearity known in biology.

| Type | Mechanism | Extent | Taxa |
|---|---|---|---|
| None | Colinear transcription | — | Animals, most fungi |
| C-to-U substitution | PPR-protein-guided deamination | ~300–500 sites per angiosperm mitogenome | Land plants (all major groups) |
| U-to-C substitution | Reverse reaction, PPR-guided | Hundreds of sites | Ferns, hornworts, lycophytes |
| U-insertion / deletion | Guide RNAs from minicircles direct the editosome to insert or delete uridines | Up to 50% of the mature mRNA sequence; some genes ("pan-edited cryptogenes" such as cox3 in T. brucei) are unreadable from DNA alone | Kinetoplastids |
| 3' oligo-U addition | Post-transcriptional U-tailing plus substitutional editing | Extensive | Dinoflagellates |
| Nucleotide insertion (mono/dinucleotide) | Co-transcriptional insertion of C, U or dinucleotides not templated by DNA | ~1 insertion every 25 nt | Physarum polycephalum and other myxomycetes |
| tRNA 5' editing | Acceptor-stem mismatch repair | Most tRNAs | Acanthamoeba, Spizellomyces, some chytrids |
| Polyadenylation-generated stop codons | Incomplete T/TA completed to UAA by poly(A) addition | Several mRNAs | Vertebrates |
| Trans-splicing of gene modules | Genes fragmented across chromosomes and joined at RNA level | Genome-wide | Diplonema, Euglenozoa; nad1/2/5 in plants |

*Table 9. Mitochondrial RNA editing and processing systems.*

Two implications are worth flagging. First, in edited systems, comparative sequence analysis of mtDNA at face value is systematically misleading — editing sites are frequently restored by editing to the ancestral amino acid, so DNA-level divergence overstates protein-level divergence. Second, editing systems are metabolically expensive and appear to be maintained largely by constructive neutral evolution rather than by any adaptive benefit; the kinetoplastid editosome is the standard example.

## 4.11 Repeats, duplications and non-coding content

Coding density spans nearly the entire possible range. Animal mitogenomes are typically 90–95% coding; the human genome has around 100 bp of intergenic sequence in total outside the D-loop. At the other extreme, less than 1% of the Cathaya argyrophylla mitogenome is coding.

- **Direct and inverted repeats** in plant mitogenomes drive high-frequency homologous recombination, generating a dynamic population of subgenomic circles rather than one stable molecule. Large repeats (>1 kb) recombine constantly; intermediate repeats recombine occasionally, producing sublimons at low stoichiometry that can be amplified by selection — the basis of cytoplasmic male sterility in maize (T-urf13), petunia (pcf) and sunflower.
- **Plastid-derived insertions (MTPTs)** are ubiquitous in plants and correlate strongly with mitogenome size above a threshold of roughly 1 Mb.
- **Transposable elements** — LINEs and LTR retrotransposons — show elevated activity in the super-large Pinaceae mitogenomes, though repeat content alone does not account for their size.
- **GC clusters** punctuate the otherwise AT-rich Saccharomyces mitogenome and mark recombination hotspots.
- **Tandem repeats in animal control regions** generate heteroplasmic length variants; in some fish and reptiles the array expands and contracts within a single individual's lifetime.

## 4.12 Copy number, heteroplasmy and inheritance

Copy number per cell ranges from a few hundred in resting lymphocytes to over 100,000 in mature oocytes, with 1,000–10,000 typical for post-mitotic tissue. Copy number is regulated separately from cell-cycle DNA replication and is often reduced in ageing tissue and in mtDNA-depletion syndromes.

- **Maternal inheritance** is the default in animals, enforced by ubiquitin-mediated and autophagic destruction of sperm mitochondria after fertilisation, plus dilution against the oocyte's enormous copy number. Reports of paternal leakage in humans exist but remain contested, with nuclear-mitochondrial segment (NUMT) artefacts a serious confounder.
- **Doubly uniparental inheritance (DUI)** occurs in many bivalve molluscs (Mytilus, Ruditapes, Venerupis). Two distinct mitogenomes coexist in a species: the F genome transmitted through eggs to all offspring, and the M genome transmitted through sperm to sons only, where it dominates the gonad. F and M genomes can differ by more than 20% at nucleotide level and carry lineage-specific ORFs. This is the only well-established stable exception to uniparental inheritance in animals.
- **Biparental and paternal inheritance** occur in some yeasts (where mating produces a recombining heteroplasmic zygote), in some conifers (paternal transmission of mtDNA in Sequoia and some Pinaceae), and sporadically in angiosperms.
- **Heteroplasmy and the threshold effect.** Pathogenic mtDNA variants typically produce no phenotype until they exceed 60–90% of molecules in a tissue. The germline bottleneck — a drastic reduction in transmitted mtDNA copy number during oogenesis — causes rapid drift in heteroplasmy level between generations and is the reason mitochondrial disease inheritance is so hard to counsel.
- **Somatic mutation and clonal expansion.** Individual cells in aged tissue accumulate clonally expanded mtDNA deletions, most prominently the 4,977 bp "common deletion" flanked by 13 bp direct repeats, which removes seven genes. In human substantia nigra and skeletal muscle, individual cells can reach near-homoplasmy for a deletion, producing the mosaic COX-negative fibres characteristic of aged muscle.

## 4.13 Nuclear and inter-organellar transfer

mtDNA is continually escaping into the nucleus. Nuclear mitochondrial DNA segments (NUMTs) are present in essentially every sequenced eukaryotic nuclear genome — the human genome contains several hundred, the largest exceeding 14 kb, and new integrations occur at appreciable rates, occasionally causing disease by insertional mutagenesis. NUMTs are the single most common source of error in mtDNA studies: they amplify with mitochondrial primers, they look like heteroplasmy, and in ancient-DNA and low-input work they can dominate the signal.

Transfer runs in other directions too. Plant mitogenomes routinely import plastid DNA (MTPTs) and nuclear DNA. Amborella trichopoda has horizontally acquired the equivalent of six entire foreign mitochondrial genomes from mosses, green algae and other angiosperms, apparently via epiphyte contact and fusion. Horizontal transfer of individual genes and introns between plant species is common enough that mitochondrial gene trees in plants must be interpreted with caution.

## 4.14 Cryptic ORFs and mitochondrial-derived peptides

The annotation of 13 protein genes in human mtDNA is now understood to be incomplete. Short open reading frames nested within the rRNA genes and elsewhere encode small bioactive peptides:

- **Humanin** — a 24-amino-acid peptide encoded within MT-RNR2, cytoprotective and anti-apoptotic, with circulating levels that decline with age.
- **MOTS-c** — a 16-amino-acid peptide encoded within MT-RNR1, which translocates to the nucleus under metabolic stress and regulates nuclear gene expression via AMPK; effectively a retrograde signal encoded in the mitochondrial genome itself.
- **SHLP1–6** — six small humanin-like peptides, also from MT-RNR2, with varying effects on apoptosis and metabolism.
- **ymf and orf genes** — fungal and protist mitogenomes carry numerous unassigned reading frames, some clearly functional and conserved, many still unannotated. Saccharomyces var1 (a ribosomal protein) was one such orphan until characterised.

The methodological point is that mitogenome annotation pipelines built on the vertebrate 37-gene template will systematically miss this class, and published gene counts for non-model organisms should be treated as lower bounds.

# 5. Survey by lineage

The same variation, re-cut by taxon. Sizes are typical values for the group; where a single species is named the figure is for that species.

## 5.1 Metazoa — vertebrates

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Mammals | 16–17 kb circular | 13 protein, 22 tRNA, 2 rRNA | Highly conserved gene order; control region with VNTRs in carnivores and ungulates; code 2 |
| Birds | 16–19 kb circular | 37 | tRNA-Ile/Gln/Met rearrangement relative to other vertebrates; control-region duplication in parrots, cuckoos, some passerines |
| Squamates (lizards, snakes) | 16–20 kb circular | 37 | Duplicated control region in boas and vipers; frequent tRNA translocations by duplication-random loss |
| Crocodilians and turtles | 16–18 kb circular | 37 | Conserved; turtles show occasional control-region duplication |
| Amphibians | 16–22 kb circular | 37 | Extensive tRNA rearrangement in some frogs (Neobatrachia); large repeat arrays in salamanders |
| Teleost fishes | 16–18 kb circular | 37 | Occasional control-region translocation and tRNA shuffling; extensive population-level use of the control region |
| Chondrichthyans, agnathans | 16–20 kb circular | 37 | Conserved; lamprey and hagfish mitogenomes retain the standard set |

*Table 10. Vertebrate mitochondrial genomes.*

## 5.2 Metazoa — invertebrates

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Tunicates (Ciona, Halocynthia) | ~14–15 kb circular | 37 (13 protein) | Genetic code 13 (AGA/AGG = Gly); highly rearranged; all genes on one strand in Ciona |
| Cephalochordates (Branchiostoma) | ~15 kb circular | 37 | Near-vertebrate architecture; code 5 |
| Hemichordates — Pterobranchia | ~15–16 kb circular | 37 | Codes 24 and 33, unique to this group (AGG = Lys) |
| Echinoderms | ~15–16 kb circular | 37 | Code 9 (AAA = Asn); large inversion distinguishing sea urchins from starfish |
| Insects (Drosophila, Lepidoptera) | 15–20 kb circular | 37 | Large AT-rich control region; code 5; gene order highly conserved except Hymenoptera and Thysanoptera |
| Sucking lice (Pediculus, Phthiraptera) | ~18 minicircles, 3–4 kb each | 37 distributed across minicircles | The most fragmented animal mitogenome; each minicircle carries 1–3 genes plus a shared non-coding region |
| Crustaceans | 15–17 kb circular | 37 | Ancestral pancrustacean order; Armadillidium vulgare has coexisting linear monomer and circular dimer |

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Chelicerates (spiders, mites, ticks) | 14–17 kb circular | 37, sometimes fewer tRNAs | Severely truncated tRNAs lacking arms; some ticks have duplicated control regions |
| Nematodes | 12–26 kb circular; multipartite in Globodera | 12 protein (atp8 lost), 22 tRNA, 2 rRNA | Armless and arm-deficient tRNAs; split LSU rRNA; code 5 or 14; Globodera has a multipartite genome |
| Platyhelminthes (flukes, tapeworms) | 13–16 kb circular | 12 protein (atp8 lost), 22 tRNA, 2 rRNA | Codes 9, 14, 21; all genes on one strand; large non-coding repeat region |
| Annelids | 15–16 kb circular | 37 | Conserved order within clitellates; all genes on one strand |
| Molluscs — gastropods and cephalopods | 14–20 kb circular | 37 | Hypervariable gene order; truncated tRNAs |
| Molluscs — bivalves | 16–32 kb circular | 36–37 (atp8 often absent) | Doubly uniparental inheritance with distinct F and M genomes; extreme gene-order variability; supernumerary ORFs in the M genome |
| Brachiopods, bryozoans | 14–16 kb circular | 37 or fewer | Some bryozoans have lost several tRNAs |
| Rotifers (Brachionus) | 2 circular chromosomes, ~11 + ~12 kb | 37 split across both | Bipartite genome with a shared non-coding region |
| Cnidaria — Anthozoa | 16–21 kb circular | 13 protein, 2 rRNA, only trnM and trnW | Group I introns in cox1 and nad5 carrying homing endonucleases; code 4; very slow mutation rate |
| Cnidaria — Medusozoa | 15–22 kb linear, 1–8 chromosomes | 13 protein, 2 rRNA, 1–2 tRNA | Linear with inverted-repeat telomeres; polB gene of plasmid origin at the termini; Henneguya salminicola has lost mtDNA entirely |
| Ctenophora (Mnemiopsis, Pleurobrachia) | ~10 kb circular | 13 protein, 2 heavily truncated rRNA, 0 tRNA | Smallest and fastest-evolving animal mitogenome; no tRNAs; loss of atp6 in some |
| Placozoa (Trichoplax) | ~43 kb circular | 13 protein + atp9, 24 tRNA, 2 rRNA, extra ORFs | Largest animal mitogenome; group I introns; fragmented cox1; large intergenic tracts |
| Porifera (sponges) | 16–25 kb circular | 13 protein + atp9, 24–27 tRNA, 2 rRNA | Most bacteria-like animal mitogenomes; retain atp9 and extra tRNAs; group I introns in cox1; standard-like code in some |

*Table 11. Invertebrate mitochondrial genomes.*

## 5.3 Fungi

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Saccharomyces cerevisiae | ~85.8 kb circular-mapping | 8 protein (cox1–3, cob, atp6, atp8, atp9, var1), 24 tRNA, 2 rRNA | ~80% AT; complex I entirely absent; group I and II introns; ori/rep elements; code 3; petite mutants readily lose most of the genome |
| Schizosaccharomyces pombe | ~19.4 kb circular | 8 protein, 25 tRNA, 2 rRNA | Compact; few introns; also lacks complex I |
| Candida, Pichia (some) | 20–50 kb linear or circular | 8–14 protein | Linear genomes with telomeric repeats in several species; code 3 or 4 |

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Neurospora crassa | ~64.8 kb circular | 14 protein (complex I retained), 27 tRNA, 2 rRNA | Group I intron-rich; code 4; senescence-associated plasmid insertions |
| Podospora anserina | ~100 kb circular | ~14 protein | senDNA plasmids excise and amplify, causing a heritable senescence syndrome — a direct link between mitogenomic instability and organismal ageing |
| Yarrowia lipolytica | ~48 kb circular | 14 protein incl. full nad set | Retains complex I; widely used as a yeast complex I model |
| Chytridiomycota (Spizellomyces) | ~60 kb, some linear | 14 protein | tRNA 5'-end editing; some linear genomes |
| Microsporidia | Absent | — | Mitosomes only; complete loss of the mitochondrial genome and of oxidative phosphorylation |

*Table 12. Fungal mitochondrial genomes.*

## 5.4 Viridiplantae

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Angiosperms (typical) | 200 kb – 1 Mb; multiple recombining circles | ~30–40 protein (full nad set, atp1/4/6/8/9, cox1–3, cob, ccmB/C/F, matR, mttB, sdh3/4), ~15–20 rps/rpl, 3 rRNA (18S, 26S, 5S), 15–25 tRNA plus imported | Standard code; ~300–500 C-to-U editing sites; ~20–26 group II introns, several trans-spliced; repeat-mediated recombination; plastid and nuclear insertions |
| Brassica napus | ~222 kb | Standard angiosperm set | Among the most compact angiosperm mitogenomes |
| Arabidopsis thaliana | ~367 kb | Standard set | The reference plant mitogenome |
| Zea mays | ~570 kb | Standard set | Master circle plus subgenomic isoforms; T-urf13 chimeric ORF causes cytoplasmic male sterility |
| Cucumis melo | ~2.9 Mb, 3 chromosomes | Standard set | Expansion by short-repeat proliferation |
| Amborella trichopoda | ~3.9 Mb | Standard set plus foreign genes | Six genome-equivalents of horizontally acquired foreign mtDNA from mosses, algae and other angiosperms |
| Silene noctiflora / S. conica | 6.7 Mb / 11.3 Mb, ~60 chromosomes | Standard set, some genes lost | Extreme multichromosomal expansion; most sequence of unknown origin; accelerated substitution rates |
| Viscum scurruloideum | ~66 kb | All nad genes lost; matR, ccm genes lost | Smallest angiosperm mitogenome; no functional complex I |
| Gymnosperms — Pinaceae | 1–19 Mb | Standard set | Cathaya argyrophylla at 18.99 Mb is the largest known mitogenome; Larix sibirica ~11.7 Mb |
| Cycas taitungensis | ~415 kb | Standard set | High repeat coverage relative to size |
| Bryophytes (Marchantia, Physcomitrella) | ~105–186 kb | Gene-rich; retains rps/rpl set | Marchantia polymorpha was the first plant mitogenome sequenced (1992); little to no editing in Marchantia |

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Ferns, hornworts, lycophytes | ~200–800 kb | Standard set | U-to-C editing in addition to C-to-U — unique among plants |
| Chlamydomonas reinhardtii | 15.8 kb linear | 8 genes (nad2/4/5/6, cob, cox1, rtl, plus rRNA modules), 3 tRNA | rRNAs fragmented into 8 scrambled interleaved modules; telomeric inverted repeats; most tRNAs imported |
| Polytomella | ~13 kb linear | Similar to Chlamydomonas | Non-photosynthetic; linear with terminal inverted repeats |
| Scenedesmus obliquus | ~42.9 kb | Reduced set | Genetic code 22 (UCA = Stop, UAG = Leu) |
| Prasinophytes, charophytes | ~40–70 kb circular | Gene-rich, near-ancestral | Charophyte mitogenomes are closest to the land-plant ancestral state |

*Table 13. Green plant and algal mitochondrial genomes.*

## 5.5 Protists and other eukaryotes

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Jakobids (Andalucia godoyi) | ~68 kb circular | 100 genes total: 66 protein, 34 structural RNA | The most gene-rich and most bacteria-like mitogenome known; carries a bacterial-type multisubunit RNA polymerase (rpoA–D) found nowhere else, plus secY, tufA, and Shine–Dalgarno-like ribosome-binding sites and operon structure |
| Jakobids (Reclinomonas americana) | ~69 kb circular | 97 genes | The original gene-rich exemplar; standard genetic code |
| Kinetoplastids (Trypanosoma, Leishmania) | kDNA network: ~25–50 maxicircles of 20–40 kb plus 5,000–10,000 minicircles of ~1 kb | ~18 genes on maxicircles; minicircles encode only guide RNAs; 0 tRNA genes | Catenated network topology; massive U-insertion/deletion editing; pan-edited cryptogenes unreadable from DNA; all tRNAs imported; some life stages downregulate the entire organelle |
| Diplonemids (Diplonema papillatum) | ~80 separate circular chromosomes, 6–7 kb each | Each gene fragmented into modules, one per chromosome | The most fragmented mitogenome known; genes reassembled by trans-splicing plus extensive editing |
| Euglenids (Euglena gracilis) | Heterogeneous linear fragments | ~7 protein genes | Highly reduced and fragmented; no rRNA operon in the usual sense |
| Apicomplexa (Plasmodium, Toxoplasma, Babesia) | ~5.9–7 kb linear tandem repeats or circular | 3 protein (cox1, cox3, cob), rRNA in ~20 small fragments, 0 tRNA | Smallest known mitogenome; fragmented rRNA on both strands; all tRNAs imported; code 4 |
| Ciliates (Tetrahymena, Paramecium) | ~40–47 kb linear with telomeres | ~20–40 protein incl. ymf orphans, split LSU rRNA | Linear with terminal repeats; split rRNA; many lineage-specific ORFs; code 4 |
| Dinoflagellates | Large, repeat-dominated, heterogeneous | 3 protein only (cox1, cox3, cob), fragmented rRNA, 0 tRNA | Massive repetitive DNA; trans-splicing; oligo-U 3' addition and substitutional editing; gene copies in many fragmentary versions |
| Stramenopiles (diatoms, oomycetes, brown algae) | ~35–60 kb circular | ~35 protein incl. rps/rpl set | Relatively gene-rich and conservative; Thraustochytrium uses code 23 |
| Amoebozoa (Acanthamoeba castellanii) | ~41.6 kb circular | ~40 protein incl. rps/rpl | Standard genetic code; tRNA 5'-end editing |

| Group | Size and form | Gene complement | Distinctive features |
|---|---|---|---|
| Myxomycetes (Physarum polycephalum) | ~63 kb circular | ~14 protein | Insertional RNA editing — non-templated C and dinucleotide insertions roughly every 25 nucleotides |
| Red algae (Porphyra, Cyanidioschyzon) | ~26–37 kb circular | ~25 protein | Gene-rich and compact; standard code |
| Glaucophytes, cryptophytes, haptophytes | ~30–50 kb circular | ~25–40 protein | Conservative, gene-rich, relatively little studied |
| Anaerobic reduced lineages | Absent | — | Hydrogenosomes (Trichomonas vaginalis, some chytrids and ciliates) and mitosomes (Entamoeba histolytica, Giardia intestinalis, Cryptosporidium, Microsporidia) retain the organelle without any DNA |
| Monocercomonoides sp. PA203 | No organelle | — | The only eukaryote known to have lost the mitochondrion entirely, including all mitochondrial protein-targeting machinery |

*Table 14. Protist and other eukaryotic mitochondrial genomes.*

## 6. Record-holders and extremes

| Category | Holder | Value |
|---|---|---|
| Largest mitogenome | Cathaya argyrophylla (Pinaceae) | 18.99 Mb |
| Largest before 2024 | Larix sibirica | ~11.7 Mb |
| Largest angiosperm | Silene conica | ~11.3 Mb |
| Smallest mitogenome | Plasmodium falciparum | ~5.97 kb |
| Smallest animal mitogenome | Mnemiopsis leidyi (ctenophore) | ~10 kb |
| Smallest angiosperm mitogenome | Viscum scurruloideum | ~66 kb |
| Largest animal mitogenome | Trichoplax adhaerens (Placozoa) | ~43 kb |
| Most genes | Andalucia godoyi (jakobid) | 100 genes — 66 protein, 34 structural RNA |
| Fewest protein genes (genome retained) | Plasmodium, dinoflagellates | 3 (cox1, cox3, cob) |
| Most tRNA genes | Jakobids; Larix sibirica | ~34 |
| Fewest tRNA genes (genome retained) | Plasmodium, ctenophores, dinoflagellates, kinetoplastids | 0 |
| Most fragmented genome | Diplonema papillatum | ~80 chromosomes, genes split into modules |
| Most fragmented animal genome | Pediculus humanus (human louse) | ~18 minicircles |
| Most extensive RNA editing | Trypanosoma brucei | Up to ~50% of mature mRNA sequence added post-transcriptionally |
| Most AT-biased | Plasmodium; Saccharomyces | >80% AT |
| First mitogenome sequenced | Homo sapiens (Anderson et al., 1981) | 16,569 bp |
| First plant mitogenome sequenced | Marchantia polymorpha (1992) | ~186 kb |
| Only animal known to lack mtDNA | Henneguya salminicola (Myxozoa, Cnidaria) | — |
| Only eukaryote known to lack mitochondria | Monocercomonoides sp. PA203 | — |

*Table 15. Extremes of mitochondrial genome architecture.*

## 7. Points of contact with ageing biology

Cross-species mitogenomic variation is an underused resource in ageing research, largely because comparative mitogenomics and biogerontology developed in separate literatures. Several intersections are worth setting out explicitly.

### 7.1 The D-loop as an ageing-relevant locus

The control region accumulates somatic point mutations with age at rates far exceeding the coding region, and specific control-region variants (the T414G transversion in fibroblasts, the A189G and T408A substitutions in skeletal muscle) show strong age-dependent clonal expansion. Because the D-loop contains the origin of replication and both promoters, a mutation there can alter copy number and transcriptional output without changing any protein. This makes the region a plausible site at which

mtDNA damage converts into altered mitochondrial output — relevant to any framework in which mitochondrial functional decline is upstream of nuclear metabolic signalling rather than merely a consequence of it.

### 7.2 Deletions, direct repeats and the architecture that permits them

The 4,977 bp common deletion arises by recombination or replication slippage between 13 bp direct repeats and removes ND3, ND4, ND4L, ND5, COX3, ATP6, ATP8 and five tRNAs. Its existence is a direct consequence of vertebrate mitogenome architecture: a compact circle with occasional short direct repeats and no efficient double-strand-break repair. Species differ in the number and spacing of such repeats, which offers a testable comparative prediction — deletion burden in aged tissue should track direct-repeat content, not simply metabolic rate.

### 7.3 Rate variation and lifespan

Mitochondrial substitution rate varies by more than two orders of magnitude across animals. Anthozoan corals have among the slowest mitochondrial mutation rates known, roughly 10–100 times slower than typical bilaterians, and include some of the longest-lived animals. Ctenophores sit at the opposite extreme on both axes. The correlation is suggestive but heavily confounded by body size, generation time and metabolic rate; the naked mole rat and bat literatures show that membrane composition and proton leak may matter more than mutation rate per se. This remains an open comparative question rather than a settled result.

### 7.4 Complex I loss as a natural knockout

Saccharomyces cerevisiae and Viscum scurruloideum have independently lost the entire mitochondrially encoded complex I. Both are viable. Complex I is the principal site of reverse-electron-transport superoxide production and a major determinant of the NAD+/NADH ratio, so these lineages are natural experiments in what an organism looks like without it. Yeast in particular is used as a complex-I-free control, though its loss of the whole nad gene set means it cannot model partial complex I dysfunction.

### 7.5 Mitochondrially encoded signalling peptides

MOTS-c and humanin, both encoded within the rRNA genes, are the clearest demonstration that the mitochondrial genome directly encodes retrograde signals to the nucleus rather than only respiratory-chain components. MOTS-c translocates to the nucleus under metabolic stress and modulates nuclear transcription; humanin levels decline with age and correlate with insulin sensitivity. Any model of mitochondrial-to-nuclear signalling in ageing has to accommodate the fact that some of the signal is peptide, mitochondrially encoded, and inducible.

### 7.6 Mitonuclear coevolution and the constraints it imposes

Every OXPHOS complex except complex II is a chimera of mitochondrially and nuclearly encoded subunits, assembled at a ratio of roughly 1:4. This forces tight coevolution: mitochondrial mutation rates 10–50 times the nuclear rate mean the nuclear genome is perpetually adapting to a faster-moving partner. Mitonuclear mismatch produces measurable fitness and lifespan effects in Drosophila and mouse conplastic strains, and is one of the better-supported mechanistic accounts of why hybrid dysfunction and, arguably, some age-related decline occur. It also means results from cybrid and conplastic models must be interpreted against the specific nuclear background used.

## 8. Pitfalls in comparative mitogenomics

- **NUMT contamination.** Nuclear copies amplify with mitochondrial primers and masquerade as heteroplasmy or as divergent haplotypes. Long-range PCR from purified mtDNA, or read-depth filtering in shotgun data, is the usual mitigation. This is the single most common source of published error in the field.
- **Circularity as an assembly artefact.** Assemblers are configured to circularise. A reported circular genome is weak evidence for a circular molecule in vivo, particularly in plants and fungi.
- **Plant "genome size" is ill-defined.** A plant mitogenome exists as a dynamic population of recombination isoforms. The reported size depends on whether one counts a master circle, the sum of contigs, or the non-redundant sequence content. Comparisons across studies using different conventions are unreliable.
- **Annotation-template bias.** Pipelines trained on the vertebrate 37-gene set miss orphan ORFs, fragmented genes, unusual tRNAs and split rRNAs. Gene counts for non-model organisms should be read as lower bounds.
- **Editing breaks colinearity.** In plants, kinetoplastids, dinoflagellates and myxomycetes, the DNA sequence does not specify the protein. Phylogenetic and functional inference from unedited DNA sequence is systematically wrong in these groups.
- **Control-region duplication.** Aligning the two copies in birds and snakes as though they were orthologous has produced published phylogenetic errors.
- **Length heteroplasmy.** Tandem repeats in control regions cause polymerase slippage, mixed sequence traces and false variant calls.

## 9. Nomenclature conventions

Gene naming is inconsistent across the literature, which causes real confusion when comparing across taxa:

- **Human and vertebrate genes** use the MT- prefix and uppercase: MT-ND1, MT-CO1, MT-CYB, MT-RNR1, MT-TL1.
- **Invertebrate and general animal literature** uses lowercase or mixed forms: nad1, cox1, cob, rrnS, rrnL, trnL(UUR).
- **Plant and protist literature** uses lowercase italic: nad1, cox1, cob, atp6, rps3, rrn18.
- **Fungal literature** often uses historical names: COB or cytb for cob, OXI1/OXI2/OXI3 in older yeast papers for cox genes, OLI2 for atp6.
- **COI, CO1, COX1 and cox1** all denote the same gene. The DNA-barcoding literature almost universally writes COI.

When compiling across sources, normalise to one convention early; the alternative is silent duplication of genes in presence/absence matrices.

## 10. Summary

The mitochondrial genome is not a conserved organellar fixture but one of the most evolutionarily labile genetic systems known. Every feature of the vertebrate arrangement — its size, its circularity, its gene count, its intronlessness, its 22 tRNAs, its single D-loop, its code, its colinearity with the protein it

specifies, its maternal transmission, and its very existence — is contingent, and each has been altered or abolished in one lineage or another. The invariant residue is remarkably thin: a cytochrome b gene, one or more cytochrome oxidase subunits, and a ribosome to translate them, all in service of keeping the redox-sensing components of the respiratory chain physically co-located with the membrane they serve.

For comparative work this has a methodological consequence and a conceptual one. Methodologically, no assumption carried over from the human mitogenome should be applied to another lineage without checking. Conceptually, the pattern of what is retained across three billion years and thousands of independent reduction events is itself the strongest available evidence about why any mitochondrial genome exists at all.

## 11. Verification sources and key references

The figures in this document are compiled from the comparative-mitogenomics literature and are approximate. For authoritative and current values, consult:

- **NCBI Organelle Genome Resources** — ncbi.nlm.nih.gov/genome/organelle/ — complete mitogenomes with standardised annotation.
- **NCBI Genetic Codes** — the authoritative list of translation tables, updated as new reassignments are described.
- **MitoFish, MitoZoa, GOBASE** — taxon-specific curated mitogenome databases.
- **MITOMAP** — the human mtDNA variant and disease database, and the source of rCRS numbering.

Primary literature underpinning specific claims above:


- **Anderson S. et al. (1981)** Sequence and organization of the human mitochondrial genome. Nature 290:457–465.
- **Ojala D., Montoya J., Attardi G. (1981)** tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474.
- **Lang B.F. et al. (1997)** An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493–497. [Reclinomonas]
- **Burger G., Gray M.W., Forget L., Lang B.F. (2013)** Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol Evol 5:418–438.
- **Gray M.W. et al. (2020)** The draft nuclear genome sequence and predicted mitochondrial proteome of Andalucia godoyi. BMC Biology 18:22.
- **Sloan D.B. et al. (2012)** Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria. PLoS Biology 10:e1001241. [Silene]
- **Rice D.W. et al. (2013)** Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science 342:1468–1473.
- **Skippington E. et al. (2015)** Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. PNAS 112:E3515–24.
- **Putintseva Y.A. et al. (2020)** Siberian larch mitochondrial genome... currently the largest known mitogenome. BMC Genomics 21:654.
- **Huang K. et al. (2024/2025)** The mitochondrial genome of Cathaya argyrophylla reaches 18.99 Mb. Frontiers in Plant Science 16:1556332.

- **Yahalomi D. et al. (2020)** A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. PNAS 117:5358–5363.
- **Karnkowska A. et al. (2016)** A eukaryote without a mitochondrial organelle. Current Biology 26:1274–1284. [Monocercomonoides]
- **Vlcek C., Marande W., Teijeiro S., Lukes J., Burger G. (2011)** Systematically fragmented genes in a multipartite mitochondrial genome. Nucleic Acids Research 39:979–988. [Diplonema]
- **Shao R., Kirkness E.F., Barker S.C. (2009)** The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse. Genome Research 19:904–912.
- **Jühling F. et al. (2012)** Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures. Nucleic Acids Research 40:2833–2845.
- **Breton S., Stewart D.T. (2015)** Atypical mitochondrial inheritance patterns in eukaryotes. Genome 58:423–431. [DUI]
- **Lee C. et al. (2015)** The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis. Cell Metabolism 21:443–454.