Title

Comparative mitogenomics of kingdom Fungi — evolutionary insights and metagenomic applications

Journal

Nucleic Acids Research, 54(2)

ISSN

0305-1048

Authors

Ahrendt, Steven R

Haridas, Sajeet

Stong, Spencer

et al.

Publication Date

2026-01-14

DOI

10.1093/nar/gkaf1419

Nucleic Acids Research, 2026, **54**, gkaf1419
https://doi.org/10.1093/nar/gkaf1419
Genomics

Comparative mitogenomics of kingdom Fungi – evolutionary insights and metagenomic applications

Steven R Ahrendt[1,2], Sajeet Haridas[1], Spencer Stong[1,3], Asaf Salamov[1], Andrei Steindorff[1], Kurt LaButti[1], Robert Riley[1], Igor Shabalov[1], Igor Lukashin[1], Serge Dusheyko[1], Frederik Schulz[1], Miguel F Romero[1], Juan C Villada[1], Igor V. Grigoriev[1,2,3,*], Stephen J Mondo[1,2,4,*]

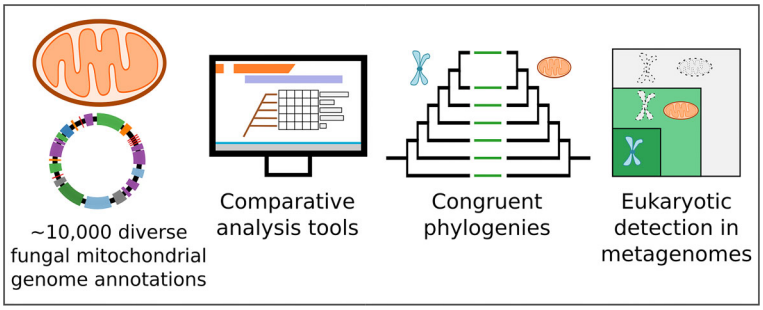
[1]U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
[3]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, United States
[4]Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80521, United States

*To whom correspondence should be addressed. Email: sjmondo@lbl.gov
Correspondence may also be addressed to Igor V. Grigoriev. Email: ivgrigoriev@lbl.gov

Abstract

Mitochondria are essential components of eukaryotic cells, responsible for ATP production through oxidative phosphorylation. Despite their biological importance, unique challenges have hindered the adoption of automated mitochondrial genome (mitogenome) annotation methods, obstructing mitochondrial comparative genomics in a broad evolutionary context. Using Fungi as a study system and a Joint Genome Institute (JGI) annotated high-quality reference set, we observed broad patterns of mitochondrial evolution across the kingdom. We found that the median fungal mitogenome size is 58 kb and identified exceptionally large examples over 1 Mb in Pezizomycetes. All 14 expected oxidative phosphorylation protein-coding genes, plus *rps3*, were generally conserved. We found evidence of major evolutionary transitions within the Ascomycota, including the transfer of mitochondrially encoded *atp8* and *atp9* to the nuclear genomes across the Pezizomycotina and shifts in mitogenome tRNA patterns across the kingdom. We found substantial concordance between mitochondrial and nuclear evolution, enabling us to document 3131 total fungal mitogenomes from JGI-derived metagenomic datasets. We also identified 6467 total undeclared mitogenomes embedded in Genbank fungal nuclear assemblies. We provide interactive tools for mitogenome analysis through the JGI MycoCosm platform. Collectively, this work generated nearly 10 000 new fungal mitogenome annotations, providing a foundation and resources for future exploration of comparative fungal mitogenomics.

Graphical abstract



Introduction

Mitochondria are specialized organelles found within cells across virtually all eukaryotic lineages [1]. They are one of the defining features of Eukaryota with a primary core function of ATP production through oxidative phosphorylation [2]. It is widely accepted that mitochondria are derived from an ancient endosymbiotic event involving alpha-proteobacteria, and as such possess their own "mitogenomes". As a result of > 1.5 billion years of coevolution, these genomes are reduced compared to free-living bacterial relatives, with many ancestral mitochondrial genes either lost or transferred to host nuclei [3].

Beyond their core role in energy metabolism, mitochondria are indispensable for their hosts as important contributors to a variety of other biological processes, including lipid metabolism, homeostasis, and stress responses [4]. Also, the role of mitochondria in aiding speciation by creating reproductive isolation has been proposed, although direct evidence for this is limited [5]. While there is no shortage of mitochondrial research among more commonly studied organisms, the

Received: May 20, 2024. **Revised:** October 15, 2025. **Accepted:** November 14, 2025
Published by Oxford University Press 2026.

increasing number of eukaryotic nuclear genome sequencing projects without a corresponding increase in mitogenome exploration introduces gaps in the understanding of biology and evolution across the Eukaryota.

Fungi, for example, harbor some of the most reduced mitogenomes of all eukaryotes, retaining only a set of genes related to electron transport and oxidative phosphorylation, translation, and tRNA processing [3]. Yet accurate, automated annotation of fungal mitogenomes has remained a challenging endeavor, often requiring manual curation to produce high-quality gene models. Many genes contain self-splicing introns, which not only often contain homing endonuclease genes [6] but can be as short as 10 bp, complicating the accurate identification of gene structure. Additionally, relaxed selection on fungal mitogenomes has led to the use of multiple genetic codes [7]: the Standard Code (code 1), the Yeast Mitochondrial Code (code 3), the Mold, Protozoan, and Coelenterate Mitochondrial Code, the Mycoplasma/Spiroplasma Code (code 4), and the Chlorophycean Mitochondrial Code (code 16) (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi).

Several automated annotation tools exist, for example, MFannot [8], MITOS [9], MitoZ [10], and DOGMA [11]. However, only MFannot is both supported and tuned for annotation of fungal mitochondria; the others are either unsupported as of this publication (DOGMA) or only applicable to non-fungal organisms (MitoZ, MITOS). Further complicating comparative mitogenomic efforts is their inconsistent representation in genome repositories: while sometimes complete and accessible, they may also be improperly embedded within nuclear genome assemblies and overlooked or separated as contaminants during submission and ignored.

Fortunately, reference nuclear genomes exist across nearly all phyla of the Kingdom Fungi, providing an excellent system within which mitochondrial evolution can be explored. As mitochondria are predominantly vertically inherited, their evolutionary history is largely thought to mirror their hosts [12] and therefore could be powerful tools for exploring the presence of eukaryotes within metagenomic datasets. Metagenomic studies are becoming increasingly popular as they provide exceptional insights into microbial communities within environments. Yet accurate assessments of the eukaryotic components of such communities, which are typically less abundant with more complex genomes, remain a challenging research area under active development (see [13, 14]). Mitogenomes are smaller and more abundant relative to their hosts, and as such represent an opportunity for documenting eukaryotic diversity in metagenomic datasets.

Few broad comparative studies have been conducted to test the mirrored evolution of mitochondria and hosts across larger timelines, for example, across an entire kingdom like the Fungi. Further, the existing complete and annotated fungal mitogenomic sequences are biased toward the Ascomycota and Basidiomycota phyla, limiting the taxonomic breadth that can be used to construct hidden Markov models (HMMs) and other profiles that could improve annotation quality. As such, large-scale comparative analyses have been mostly limited to specific clades within, e.g. the Saccharomycotina [15] or the genus *Aspergillus* [16]. Recently, Fonseca *et al.*, 2021 [17] characterized mitogenomes across ~800 fungal genomes, representing most of the kingdom. Their work revealed a wide range of GC content (<10% up to 68%) and genome sizes (12 to 531 kb), yetalso reported the absence of a single conserved marker across Fungi that could be used for accurate phylogeny reconstruction of the kingdom. Fungi, unlike many other eukaryotic kingdoms, regularly engage in multiple unusual reproductive processes that may alter the vertical mitochondrial transmission paradigm. Filamentous fungi in particular exhibit such processes as cross-species hyphal fusions, dikaryon formation and dikaryon-monokaryon matings, a parasexual cycle, and others. While rare, biparental inheritance and persistent heteroplasmy have been documented with fungal mitochondria [18]. To be useful as a marker of fungi within metagenomic datasets, it is essential to confirm that mitogenome evolution generally mirrors that of their hosts and that the impact of horizontal transmission is limited.

Here, we have used a workflow developed at the Joint Genome Institute (JGI) [19] to annotate a representative set of ~400 published, high-quality fungal mitochondria from the JGI MycoCosm genome portal (https://mycocosm.jgi.doe.gov/) [20]. We present a broad overview of mitogenome annotations and evolution from across the Fungal Tree of Life. In this dataset, we report newly identified huge fungal mitogenomes from the Pezizales that are each single-scaffold assemblies between 1.1 and 1.3 Mb in size (more than twice as large as previously reported; *Morchella crassipes*, 531 kb, [21]. Using our high-quality mitogenome representatives, we reveal broad kingdom-wide shifts in mitogenome tRNA usage, including, amongst other patterns, a major reduction in tRNA content within Chytridiomycota as well as a transition in Ascomycota of $tRNA^{Arg}$. Additionally, we explored mitogenome changes that are associated with fungal speciation within individual clades, where we found changes in gene order typically associated with genus boundaries.

Since our annotation pipeline could predict conserved core genes across the kingdom, we constructed a kingdom-wide mitochondrial phylogeny. We found substantial concordance between nuclear and mitochondrial phylogenies ($P < 0.01$) across the kingdom. Consequently, we surveyed over 26 000 published metagenomic datasets from the IMG genome portal (https://img.jgi.doe.gov) using mitogenomes as a tool to detect fungal hosts [22]. This approach identified 3131 mitogenomes that were likely to be from fungal hosts, some of these mitogenomes potentially representing new clades. We also investigated the presence of cryptic mitochondrial scaffolds in publicly available fungal genome assemblies by identifying and annotating 6467 undeclared mitochondrial assemblies embedded in NCBI fungal nuclear genome assemblies.

Finally, to increase visibility and encourage future use of mitochondrial data, we created a new comparative mitogenome page (https://mycocosm.jgi.doe.gov/fungal_mito_manuscript) and provided mitochondrial annotation results via individual MycoCosm web portals (e.g. https://mycocosm.jgi.doe.gov/Lacind1/organelle/mitochondria). This collective effort will dramatically expand the set of fungal mitogenome annotations and work toward a better understanding of fungal mitochondrial genome structure and evolution.

Material and methods

Sequencing and assembly

The JGI-derived mitogenomes represented in this manuscript were generated over a 10-year period of time between 2010–2020 and, as such, represent evolving sequencing technologies and assembly algorithms but relied on JGI's standard pro-

duction fungal assembly pipelines with stages dedicated to organelle assembly. The organelle assembly stages generally employ the same premise of isolating organellar sequence data for targeted assembly. Assembly methods for each genome are provided in Supplementary Table S2 and refer to the corresponding sections below.

The earliest automated mitochondrial assembly methods (assembly method "1") relied solely on assembling either a 0.75–2X (based on the average genome size of 40 Mb) or 2-million-read subsample. This was done to dilute away the nuclear genome and target only very highly represented content, such as organelles, from the input sequence data. Later versions (assembly method "2") offered refinement and robustness in isolating organellular data, using a 2 million read subsample generated with bbtools version 37.93 reformat.sh (http://sourceforge.net/projects/bbmap) (parameters: "sampleseed = 1″). The subsampled data were additionally filtered for length and quality using bbtools version 37.93 reformat.sh (parameters: "qtrim = t trimq = 5 minavgquality = 5 minlength = 101 maxns = 3″), and subsequently assembled with Velvet version 2.1.7 [23] using velvetg (parameters: "-cov_cutoff 20″). The resulting assembly was aligned to the NCBI refseq.mitochondrion database with BLAST megablast version 2.2.26 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with a minimum percent identity of 80% to identify organelle. A secondary assembly was performed with Velvet version 2.1.7 using cov_cutoff, max_coverage, and exp_cov cutoffs defined from the coverages associated with the contigs previously identified as organelle. Read pairs providing linking support between the assembled contigs are identified by aligning the original input fastq to a version of the assembled contigs with all bases masked with N, with the exception of the terminal 300 bases, with bwa version 0.7.4-r385 [24] (parameters: "mem -t 16″). The linking read pairs are used in conjunction with NCBI alignment results to refseq.mitochondrion to identify trusted organelle contigs. The main genome 18S ribosomal elements are identified by alignment to the NCBI nt database with BLAST megablast version 2.2.26, with a minimum percent identity of 80% and excluded from the list of putative organelle reads. An enriched set of organelle reads was then created from the original input fastq reads by kmer matching with bbtools version 37.93 bbduk.sh (http://sourceforge.net/projects/bbmap), using defaults, against the resulting list of organelle contigs to recruit potentially walking reads. Those that do not match the organelle contigs are output into a separate nonOrganelle fastq for downstream assembly. To leverage the AllPathsLG algorithm for assembly, simulated 1000 ± 50 bp insert long mate-pairs were generated from the organelle contigs with wgsim version 0.3.1-r13 (https://github.com/lh3/wgsim) (parameters: "-d 1000 -s 50″), and coassembled 125X of the enriched organelle matching read set with AllPathsLG release R46652 [25] to produce a final mitochondrion assembly.

Support for long read sequence data was added to the production fungal assembly pipeline, with updated algorithms suitable for long reads and improved organelle handling methodology, starting in 2016 (assembly methods "3", "4"). Input Falcon pre-assembled (pread) data was kmer profiled using bbtools kmercountexact.sh (http://sourceforge.net/projects/bbmap) with default parameters, filtered with bbtools bbnorm.sh [pigz passes = 1 bits = 16 min=(main peak ∗ 1.5) target = 9 999 999] (http://sourceforge.net/projects/bbmap), and assembled with Flye (parameters: "–pacbio-corr -g 100k –asm-coverage 100″) version 2.3.6 [26]. Genes were called on the initial Flye assembly using Prodigal (https://github.com/hyattpd/Prodigal) and examined for mitochondrial HMM models with hmmsearch [27]. Contigs containing mitochondrial HMM were masked of ribosomal content with bbtools bbduk.sh (parameters: "k = 25 mm = f kmask = N") and used to recruit Falcon-corrected pre-assembled reads ("preads") with bbtools bbduk.sh (parameters: "k = 25 mm = f mkf = 0.03 ordered ow"). Resulting preads were reassembled using Flye version 2.3.6 (parameters: "-g 100k –asm-coverage 100″). A total of three rounds of read recruitment and reassembly are conducted to provide the final assembly.

Certain default production organelle assemblies were identified as substandard (having > 1 contig) or failures (having no mitochondrial contigs). Improvement of such assemblies (assembly method "5") was carried out with iterative rounds of read recruitment to contigs with evidence of being of mitochondrial origin (GC, Coverage, BLAST/sketch, hmm) with bbtools bbduk.sh and reassembly until the assembly size plateaus or generates a circular contig. If no mitochondrial contigs could be identified, a variety of techniques were used to isolate mitochondrial sequences for assembly, including nucleotide composition filtering, subsampling, not subsampling, assembled mitochondrial identification and promotion for downstream recruitment, and non-mitochondrial contig filtering before read recruitment.

Overall, the different methods (1, 2, 3, 4, and 5) respectively produced 1, 147, 119, 46, and 30 assemblies (Supplementary Table S2).

Mitochondrial annotation

We used the mitochondrial annotation workflow, previously developed at the JGI [19], where protein-coding genes were predicted using three complementary methods: the prokaryotic *ab initio* gene finder algorithm Fgenesb (www.softberry.com), the homology-based protmap (www.softberry.com), and Genewise with fungal mitochondrial HMMs. The Fgenesb predictions are checked using blast against a database of known mitochondrial proteins, and both protmap and Genewise are alignment-based methods. To build HMM profiles for Genewise, we collected the non-redundant fungal mitochondrial proteins from GenBank as well as the U. Montreal curated collection [28](http://www.bch.umontreal.ca/People/lang/FMGP/proteins.html) corresponding to the 15 "core" fungal mitochondrial genes defined as *cox1*, *cox2*, *cox3*, *cob*, *atp6*, *atp8*, *atp9*, *nad1*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, and *rps3*). Gene models with overlapping coding sequences are filtered based on homology score or ORF length. tRNAs genes were predicted using tRNAscan-SE with the organellar option [29]. Ribosomal RNAs were predicted using Infernal (http://eddylab.org/infernal) with covariance models "LSU_rRNA_bacteria" and "SSU_rRNA_bacteria" from the RFAM database [30]. Annotated mitogenomes were visualized with Circos [31]. For comparison, we also ran the publicly available mitochondrial genome annotation tool MFannot with default parameters on the genomes of our high-quality dataset.

To identify the presence of potential nuclear-encoded mitochondrial genes (*atp8* and *atp9*), nuclear genome assemblies were translated in all 6 frames to generate ORFs with a minimum length of 30 amino acids. We then used the hmm profiles described above with hmmscan (HMMER 3.3.2) and an e-value cutoff of 1e-10.

Dataset

For our phylogenetic analyses, we focused primarily on fungal mitogenome assemblies available in MycoCosm that a) have fewer than 6 scaffolds, b) have at least 12 of the 15 "core" genes commonly found in fungi, and c) were sequenced alongside now-published nuclear genomes. Together, these criteria yielded a dataset of 328 mitogenomes, representing both linear and circular topologies, which we refer to here as "high-quality." We additionally included publicly available outgroup genomes from Cryptomycota, Chytridiomycota, and Blastocladiomycota lineages ($n = 8$), non-fungal Eukaryotes ($n = 15$), and the bacterium *Rickettsia prowazekii*, for a combined dataset of 352 genomes. Specific analyses (genome size and gene order) included additional Pezizaceae ($n = 5$) and Phyllosticta ($n = 12$) genomes. The complete list of 369 species and their sources can be found in Supplementary Table S2.

Phylogeny reconstruction and clustering

To construct the mitogenome phylogeny, core genes (one copy per genome) were aligned using Muscle v3.8.1551 [32] and trimmed with Trimal v1.4.rev15 build[2013–12-17] [33] (parameters: "-automated"). The trimmed alignments were then concatenated, and the phylogeny was built with IQ-Tree version 2.1.2 [34] (parameters: "-m TEST + R11 + C60 –msub mitochondrial –fast -T AUTO -b 100").

To explore conservation of hypothetical clusters, for all 352 fungal mitogenomes, known core genes were removed, then all remaining proteins were extracted and annotated using PFAM version 34 [35]. These non-core genes were then filtered to remove any models with mobile element-related PFAM domains. We also removed any models that had PFAM domains present on known mitochondrial genes, which could represent core gene fragments. This led to the removal of any model with the following PFAM domains: COX1, COX2, COX2_TM, COX3, Cytochrom_B_C, Cytochrom_B_N_2, Cytochrome_B, GIY-YIG, LAGLIDADG_1, LAGLIDADG_2, LAGLIDADG_3, NADH5_C, NADHdh, NUMOD1, NUMOD3, Proton_antipo_M, Proton_antipo_N, RVT_1, RVT_N. Filtered hypothetical genes were then clustered using MCL [36] with inflation factor 1.5.

The fungal nuclear phylogeny was constructed by first clustering all fungal protein sequences using mmseqs2 version 0188988235c6f1a8e90f327827c73f981db8a19a (parameters: "–cov-mode = 0.5") [37]. For clusters where over 40% of lineages were present, one representative sequence per individual was selected, aligned with mafft v7.475 [38] using default parameters, then trimmed with gblocks version 0.91b [39] (parameters: "-t = p -e = .gb -b4 = 5 -b5 = h"). After trimming, the remaining 666 420 sites from 3031 aligned protein sequences were concatenated, and the phylogeny was reconstructed using FastTree version 2.1.10 SSE3 [40].

Identification of hidden genomes from public databases

We downloaded 11 491 fungal nuclear genome assemblies from GenBank on October 10, 2022. After removing 2236 assemblies with declared mitochondrial scaffolds, we used an HMM-based method to search the remaining 9255 assemblies for scaffolds with any of the 15 "core" fungal mitochondrial genes and mitochondrial homing endonuclease genes (GIY-YIG and LAGLIDADG). Scaffolds with at least one copy of any mitochondrial genes were collected and annotated using the same pipeline as described previously for the MycoCosm-derived dataset, resulting in 6467 mitogenome assemblies.

Identification of fungal mitochondria in metagenomic datasets

Our approach to identifying fungal mitochondria in metagenomic datasets involved two main steps: first, a broad search for mitochondrial rDNA sequences, and second, a more comprehensive analysis to identify and validate mitochondrial metagenome assembled genomes (MAGs).

In the initial step, we searched 26 257 public metagenomes from the IMG database for contigs containing mitochondrial rDNA sequences. We used cmsearch (infernal v1.1.4 [41]) with the bacterial SSU rRNA covariance model RF00177 from the Rfam database [42]. To include long sequences, we selected the –anytrunc option in cmsearch and assembled sequential, non-overlapping alignments into single sequences. We then annotated the recovered SSU rRNA sequences using the Silva [43], PR2 [44], and MycoCosm [20] databases with blastn (v.2.13.0), retaining those with a mitochondrial sequence as the best hit and an alignment length $\geq$ 500 bp. This process yielded 39 929 candidate mitochondrial SSU rRNA genes, of which 8743 contigs with length $\geq$ 5 Kbp were retained for subsequent analyses.

Building upon these initial findings, we then conducted a more comprehensive analysis to identify mitochondrial MAGs. We gathered 18 621 public and unrestricted samples from the IMG/M database (accessed in October 2022) that contained underlying read depth information. These samples included metagenomes, single particle sorts, and cell enrichment projects. To streamline our analysis, we focused on contigs $\geq$ 5 kbp in length. We performed metagenomic binning on each sample using Metabat2 version 2.15 [45], with specific parameters (–minContig 5000, –minClsSize 20 000, and –cvExt).

We identified putative mitochondrial MAGs by linking the contigs that contained a fungal mitochondrial SSU to the set of MAGs generated through binning. We then used CheckM2 v1.0.2 (https://github.com/chklovski/CheckM2) to estimate the contamination of the MAGs, excluding those with contamination $\geq$ 10% from further analysis. For excluded MAGs, we retained their corresponding contigs containing the putative mitochondrial 16S for subsequent analyses. We then performed gene prediction to each bin using Prodigal (-p meta) (version 2.6.3) (https://github.com/hyattpd/Prodigal).

To confirm their phylogenetic placement, we subjected fungal mitochondrial MAGs to phylogenomic analysis, comparing them with known fungal mitochondria downloaded from NCBI Genbank.

We built an initial species tree from a concatenated alignment of 17 electron transport chain genes (056_ND1, 062_ND2, 049_ND3, 061_ND4, 059_ND4L, 060_ND5, 058_NAD6, 052_NDUFS2, 051_NDUFS3, 069_CYTB, 018_COX1, 019_COX2, 020_COX3, 005_ATP5F1A, 006_ATP6, 004_ATP8, and 009_ATP5MC2) using nsgtree (https://github.com/NeLLi-team/nsgtree). These marker genes ("mitoETC") represent a set of eukaryotic genes described previously [46]. Only genomes that had at least 20% (3 out of 15) of the marker genes were included in the alignment. Alignments were created with mafft(v7.31) [38] trimmed with trimal (v1.4) [33], concate-

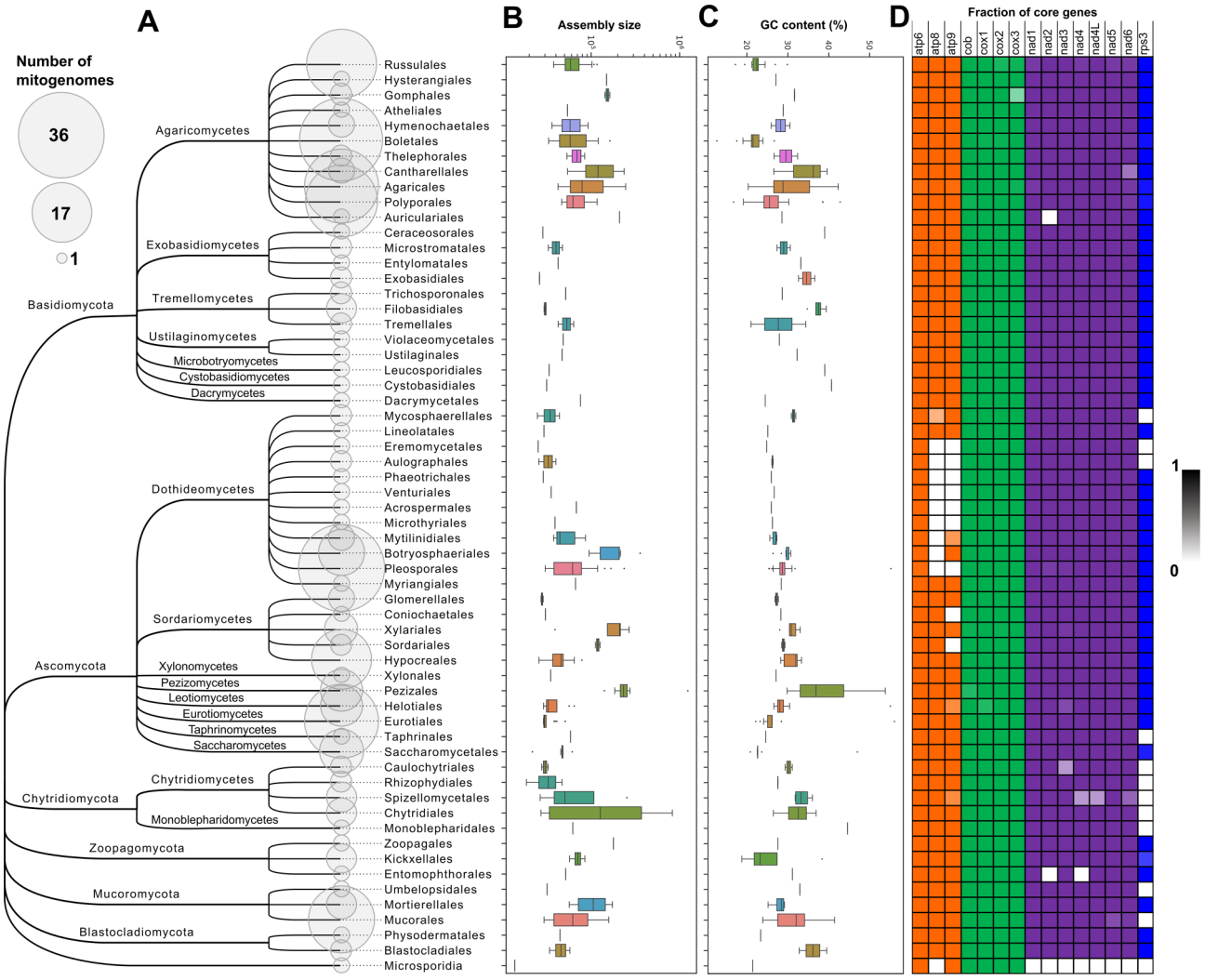


Figure 1. Features summary of 332 JGI high-quality fungal mitogenomes by lineage. (**A**) Dendrogram based on order level and built using the CommonTree tool (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). Bubbles on the tips of the tree represent the number of genomes analyzed. (**B**) Distribution of mitochondrial assembly size. The *y*-axis is represented in log10 of base pairs. (**C**) Distribution of GC content of mitochondrial assemblies. (**D**) Heatmap showing the fraction (ranging from 0 to 1) of each core gene by order level.

nated, and the species built with FastTree v2 [47]. PhyloDM (https://github.com/aaronmussig/PhyloDM) was used to infer pairwise evolutionary distances from the initial species tree.

Genomes in this initial tree were then clustered based on the resulting matrix of pairwise evolutionary distances using MCL v14-137. The most complete genomes from each cluster, defined as those with the highest count of different mitoETC proteins and the lowest number of duplications of mitoETC proteins, were used as cluster representatives. The tree was manually inspected, and MAGs that grouped monophyletically with known fungal mitochondria and the closest branching non-fungal mitochondria (Opisthokonta - Cnidaria) were collected. We built a final species tree using only the cluster representatives with IQ-tree (v2.03) and the LG4X model with 1000 ultrafast bootstrap replications. This yielded a combined set of 1791 MAGs and 1340 single contigs (3131 combined) that grouped in a monophyletic clade with known fungal mitochondria. The 3131 mitochondrial SSU rRNA gene sequences were clustered using vsearch v2.28 [48] with the –cluster_fast parameter at 97% identity.

Results

Fungal mitogenomes are diverse in size, architecture, and content

Genome size. Across our dataset, we observed substantial diversity in mitochondrial genome sizes. The median fungal mitogenome size was 58 kb, though we found size to be clade-specific (Fig. 1). Consistent with previous publications, the Cryptomycota harbor the smallest mitogenomes: *Rozella allomycis* at 12.1 kb [49] and *Mitosporidium daphniae* at 13.7 kb [50]. Conversely, the largest mitogenomes encountered are in the Ascomycete family Pezizaceae. At over 1 Mb, these are twice as large as the previously reported largest fungal mitogenome, *Morchella crassipes* at 531 kb [21]. Specifically, these large mitogenomes can be found in the ectomycorrhizal desert truffles ($n = 5$; mean = 1.2 Mb) within the genera *Terfezia* (*T. boudieri* at 1.2 Mb and *T. claveryi* at 1.2 Mb), *Mattirolomyces* (*M. terfezioides* at 1.3 Mb), *Tirmania* (*T. nivea* at 1.2 Mb), and *Kalaharituber* (*K. pfeilii* at 1.1 Mb), and are larger than the others in our dataset within the Pezizaceae ($n = 7$; mean = 0.2 Mb) (Welch two-sample, $t = 24.4$,

df = 5.6, $P < 0.0001$). Analyses of these large genomes suggest that while the average number of introns per genome (mean = 29.6) is consistent with that of others in this group (mean = 29.9), they have some of the largest introns, with cumulative intron lengths ranging from 294 492 to 405 910 nt (mean = 373 390 nt). For Pezizaceae, intron lengths range from 77 760 to 181 894 nt (mean = 113 557 nt). These large genomes also have more and longer genes encoding hypothetical proteins, with counts between 40 and 143 (1 to 17 for other Pezizaceae), and the median lengths per genome ranging from 17 670 to 46 911 nt (2151 to 7806 nt for other Pezizaceae).

Additional diversity can be observed within certain genera where our dataset contains multiple species (*Phyllosticta*, *Fusarium*, *Suillus,* and *Lactarius*). For example, mitogenome sizes in genus *Phyllosticta* correlate with lifestyle (Welch two-sample, $t = -48.2$, df = 9.3, $P < 0.0001$), showing a bimodal distribution where the pathogenic strains (n = 12) of *P. citricarpa*, *P. paracitricarpa*, *P. citriasiana,* and *P. citrichinaencis* are > 200 kb (mean 209.3 kb), while non-pathogenic strains ($n = 8$) of *P. citribraziliensis* and *P. capitalensis* (sometimes described as an opportunistic pathogen) are <150 kb (mean 125.8 kb) (Supplementary Fig. S7). The mitogenomes from the genus *Fusarium* show a general reduction in size, though no such pattern is seen in either *Suillus* or *Lactarius* genomes.

GC content. We observed similar diversity in mitochondrial genome GC content, ranging from 18.4% to 44.6% (mean 27.9%), which is distinct from the nuclear genome GC content (26.5% to 67.6%, mean 49.4%; Welch two-sample, $t = -52.3$, df = 646.3, *P*-value < 0.0001). Specific phyla have distinct GC content profiles (Supplementary Fig. S2). In most instances, phylum-level GC% relative to the average was consistent between mitochondrial and nuclear genomes. An exception was observed in the Mucoromycotina ($n = 52$), where the mean nuclear GC% of 38% was far lower than that overall (Welch two-sample, $t = -8.7$, df = 28.3, *P*-value < 0.0001). We also found unusually high GC profiles in certain lineages, for example: *Hyaloraphidium curvatum* at 43% (Chytridiomycotina), *Trametopsis cervina* at 43% (Agaricomycotina), and *Pyronema omphalodes* at 45% (Pezizomycotina). (Supplementary Fig. S2).

Gene content. Previous kingdom-wide analyses used MFannot. While MFannot in general performed comparably on our dataset to the JGI workflow, we note the following differences. The JGI workflow consistently captures more and larger known genes, with a narrower range of predicted sizes (Supplementary Fig. S1A). Additionally, MFannot failed to predict at least one core gene in 152 of the 328 JGI genomes. The most commonly missed gene was *rps3*; however, there were a small number of instances where other genes were missed (Supplementary Fig. S1B, C). Importantly, the JGI pipeline reliably finds conserved marker genes across all Fungi surveyed (Fig. 1), allowing us to construct a kingdom-wide mitochondrial phylogeny (Fig. 2). In our dataset, fungal mitogenomes contain all expected subunits of cytochrome c oxidase (Complex IV) and cytochrome bc1 (Complex III). Of the three mitogenome-encoded subunits of ATP synthase (Complex V), atp6 is consistently present, but both atp8 and atp9 are absent from at least eight orders within the class Dothideomycetes (mostly Pleosporales). However, ten species in the Botryosphaeriales, eight of which are in the genus *Phyllosticta*, are only missing atp8 and encode a copy of atp9. Searching against the nuclear genomes of our study set with hmmscan, we identified mostly single nuclear copies ("nc-") of *atp8* almost exclusively in the Dothideomycetes (95% of genomes), while atp9 was identified in 92% of the broader Pezizomycotina (including the Dothideomycetes, and found in multiple copies in some lineages, like *Chaetomium* spp.). Both of these genes are almost entirely absent from the nuclear genomes of all other clades (Supplementary Fig. S7). Our findings, which cover a wider breadth of fungi, are consistent with previous smaller analyses reporting on the transfer of atp9 [51], [52] and atp8 [53]. We similarly surveyed all 15 core genes and found that Chytridiomycota and Mucoromycota were also completely missing a mitochondrially-encoded *rps3*, which is the only ribosomal protein encoded by fungal mitochondrial genomes and is known to be absent in some early lineages [54]. We failed to detect nc-*rps3* within the nuclear genome of Chytridiomycota; however, we observed its presence in Mucoromycota. In addition to simply identifying nuclear encoded genes, we surveyed them for mitochondrial targeting signals using DeepLoc 2.0 [55] and found that the majority of nuclear nc-*atp8* and nc-*atp9* genes, as well as all nc-*rps3* genes, display mitochondrial targeting signals. While rare, we see sporadic evidence of other core genes within the nuclear genomes, some of which are similarly supported by the presence of mitochondrial targeting signals (e.g. *nad1* in *Cantharellus anzutake* and *nad4* in *Cryptococcus terricola*). As previously observed [49], we also found that NADH:ubiquinone oxidoreductase (Complex I) subunits are missing completely from Cryptomycota lineages (Fig. 1).

Across our representative mitogenome dataset, our annotation pipeline also detected up to 53 hypothetical ORFs per lineage (average 6). After filtering to remove hypothetical ORFs with homology to known core fungal mitogenes and mobile elements (see methods), MCL clustering was performed to identify gene families across the 1951 predicted hypothetical mitochondrial proteins. This resulted in 767 clusters, including 272 multi-copy clusters (Supplementary Table S1). However, most contained fewer than 10 genes. Several lines of evidence exist suggesting some of these hypothetical genes could be functional. For example, many gene clusters are found in single, but sometimes ancient, clades (e.g. cluster 1, present across Agaricomycotina; Supplementary Fig. S3), a pattern we would not expect in nonfunctional genes given the rapid mutation rate in mitochondria [56] and rapid loss of non-essential genes in endosymbionts [57]. For the 22 clusters with >10 genes, the average copy number was 1.27x, indicating that these hypothetical genes are mostly single copy within mitogenomes. Some clusters also contained PFAM domains, including the largest multicopy cluster (220 genes from 108 lineages, mostly Agaricomycotina), where 87.73% of genes bore DNA_pol_B_2 (PF03175) domains, in agreement with the recently characterized dpo distribution across fungi [17].

Phylum-level tRNA patterns detected across Fungal mitogenomes

Mitochondrial genetic codes are known to vary across phylogenetically diverse fungi [7], however, little is known about mitogenome tRNA distribution patterns. To provide a global picture of fungal tRNA evolution, we analyzed the results of tRNAscan-SE across our representative set of high-quality mitogenomes. We found that mitogenomes across the kingdom encode a median of 27 tRNA genes, except for Chytridiomycota. Within this phylum, we detected a reduced set of at most seven tRNA

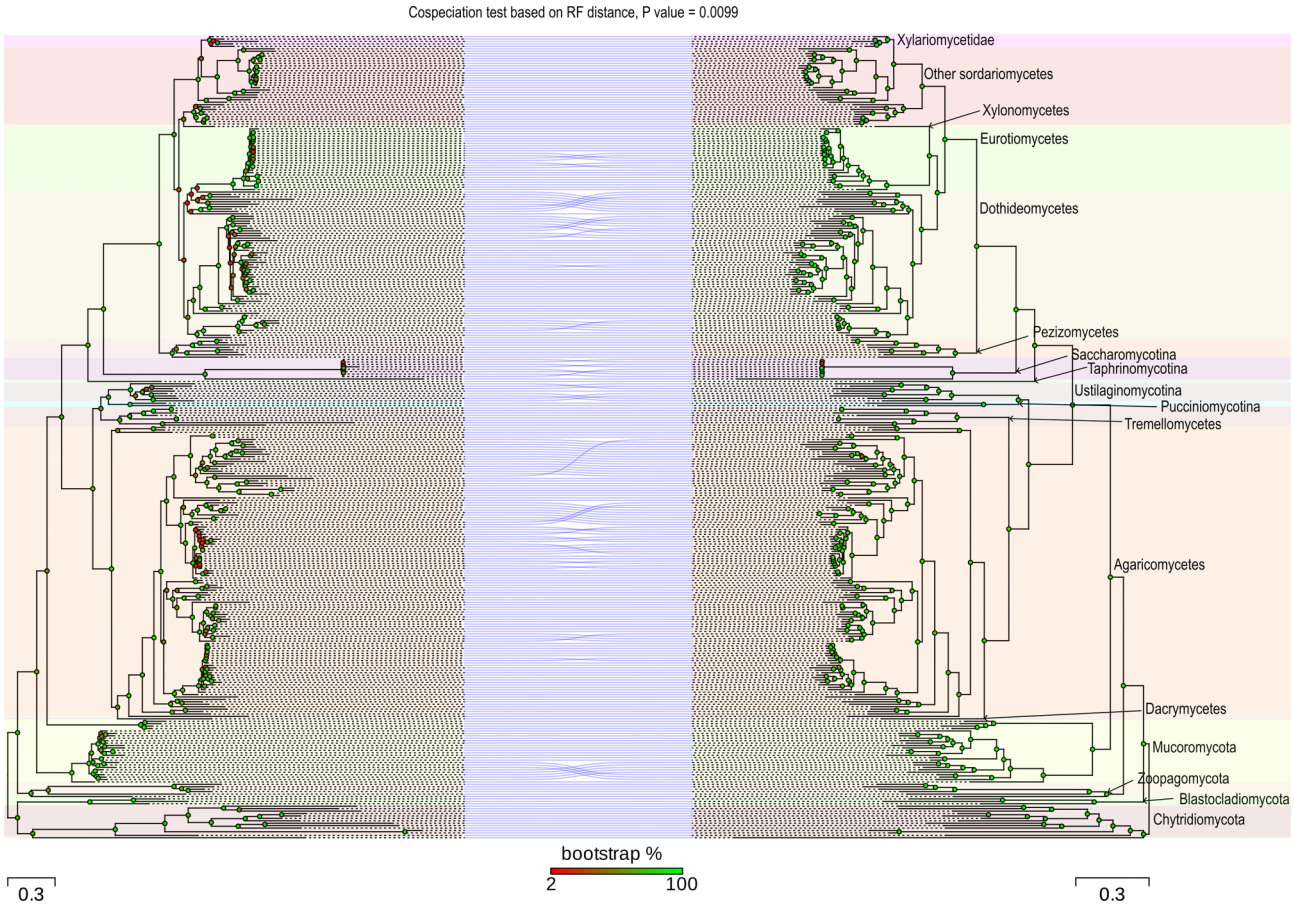


Figure 2. Cophyly between mitochondria and their hosts. Links are drawn between mitochondria (left) and hosts (right), revealing overall strong coevolution ($P = 0.0099$ using the cospeciation test, as implemented in Phytools) across the fungal kingdom. The mitochondrial tree was generated using IqTree based on protein sequences from all core mitogenes. 3031 protein orthologs were used to construct the nuclear tree, which was generated using FastTree. Bootstrap support values across both trees range from 2 to 100.

genes: $tRNA^{Lys}$(UUU), $tRNA^{Met}$(CAU), $tRNA^{Pro}$(UGG), $tRNA^{Gln}$(UUG), $tRNA^{Trp}$(CCA), $tRNA^{Tyr}$(GUA), and $tRNA^{Asp}$(GUC). Additional losses occurred within the Rhizophydiales, where $tRNA^{Asp}$(GUC) was not detected. Chytridiomycota also uniquely possess a suppressor tRNA for codon UAG. The presence of this suppressor tRNA generally corresponds to organisms that use the Chlorophycean Mitochondrial Code (like *Spizellomyces*) [58], except for Monoblepharidomycetes, which use the more common codon translation code 4. We also detected several other global trends across fungi, which we anticipate have important evolutionary and functional implications for mitochondria.

To explore discrepancies between codon and tRNA presence, we compared mitogenome tRNA presence/absence to codons available within complete mitochondrial core genes. We found that even though most codon variants are present within core genes, fungal mitochondria typically retain only one tRNA per amino acid (Supplementary Fig. S4). We also observed several cases where predicted tRNAs exist, but no corresponding codons could be detected, for example, $tRNA^{Cys}$(GCA). We found many instances (more so than any other tRNA) of a lack of the UGC codon, while still retaining the $tRNA^{Cys}$(GCA) gene. The advantage of retaining such tRNAs across evolutionary time remains unclear.

We also found evidence of shifts in tRNA utilization. These include a phylum-wide shift in Arginine tRNA usage ($tRNA^{Arg}$). Except for Chytridiomycota, which lack $tRNA^{Arg}$ genes, all Fungi have $tRNA^{Arg}$(UCU). Blastocladiomycota, "Zygomycetes", and Basidiomycota all additionally possess $tRNA^{Arg}$(UCG). However, at the base of the Ascomycota, we find a replacement of $tRNA^{Arg}$(UCU) with $tRNA^{Arg}$(ACG). Some exceptions/possible reversions appear to have occurred over time, including in *Yarrowia* (Saccharomycotina) and *Phyllosticta* (Dothideomycetes). Beyond $tRNA^{Arg}$, several other global trends were observed, including a lack of $tRNA^{Lys}$(GAG) in fungal mitogenomes, despite the frequent presence of the CTC codon. Additionally, the $tRNA^{Trp}$ gene and corresponding UGG codon are sporadically missing from members of the Russulales, and the $tRNA^{Trp}$ gene is completely absent from Ascomycota. We also found that while fungal nuclear genomes generally have $tRNA^{Ile}$(AAU) and $tRNA^{Ile}$(UAU) genes, $tRNA^{Ile}$(GAU) appears mitochondria-specific.

Finally, we can observe distinct genus-specific profiles within *Aspergillus*. We see the appearance of $tRNA^{Gly}$(ACC), which is rarely seen in other fungi and mito-specific, and not only the disappearance of the $tRNA^{Trp}$ gene seen in other Ascomycota, but additionally the corresponding TGG codon. Additionally observed is the general absence of codon GGC, which is generally present in other Fungi. Taken together, these

kingdom-wide mitogenome patterns demonstrate a shifting evolutionary landscape of tRNA usage.

Mitochondrial evolution mirrors host evolution

Since our annotation pipeline successfully detected all 15 core genes across fungi, we were able to create a kingdom-wide mitochondrial phylogeny. We compared this to a maximum likelihood fungal nuclear phylogeny, and while incongruities were observed, overall, we detected a high degree of concordance between the two phylogenies ($P = 0.009901$, Co-speciation test based on Robinson Foulds distance, 100 simulations; Fig. 2). Incongruities were often observed in locations with low bootstrap support on the mitochondrial tree, which often occurred in abundantly sampled genera (e.g. *Aspergillus*; Eurotiomycetes, *Yarrowia*; Saccharomycotina, *Suillus*; Agaricomycetes). Overall, our results demonstrate that mitochondrial and nuclear tree topologies are highly congruent, even across large phylogenetic distances.

Unidentified mitochondrial genomes in GenBank fungal nuclear assemblies

We identified 6467 fungal genome assemblies in GenBank with at least one undeclared mitochondrial scaffold. Of these, 202 (3%) are complete (all 15 core genes present), 1574 (24.4%) have more than 11 core genes, and 4893 (75.6%) have fewer than 10 core genes. Regarding the number of scaffolds, 2182 (33.7%) are single scaffolds, and the other 64.3% contained two or more scaffolds. Considering only the mitogenomes with < 6 scaffolds and 12 or more core genes yielded 957 assemblies, defined as high-quality. Sordariomycetes (310 - 32.4%) and Eurotiomycetes (397 - 41.5%) classes cover most of them, followed by Dothideomycetes (114 - 11.9%) and Saccharomycetes (54 - 5.6%). The only high-quality non-Dikarya mitogenomes found were from Mucoromycota (19 - 1.98%) (Supplementary Table S3, Supplementary Fig. S5, Fig. S6). We compared our annotations of the 957 NCBI high-quality mitogenomes with those already present in GenBank. Only 225 (23.5%) were deposited with gene predictions, and among the scaffolds predicted as mitochondrial, an average of 5.24 genes were predicted (12.32 for JGI mitochondrial annotation), with 78 (34.6%) having 0 genes. This discrepancy can be explained by the difference between Eukaryotic/mitochondrial gene prediction (e.g. codon usage, exon/intron structures). Since Eukaryotic gene predictors were used on NCBI mitochondrial scaffolds, these gene models are mostly fragmented, chimeric (a combination of two or more gene models), or simply wrong due to using the wrong genetic code. These results show that even with an untargeted mitochondrial assembly, a significant amount of high-quality mitochondrial genomes can be detected in public repositories, providing a rich dataset for further analysis.

Mitochondria are common in metagenomics datasets

From the metagenomic sequencing datasets, a total of 3131 assemblies (comprising 1340 single contigs and 1791 MAGs) were found with our phylogenomic approach to likely represent partial or complete fungal mitochondria (Fig. 3). Collectively, 1867 were in single scaffolds with 2994 in five or fewer scaffolds. A total of 388 were determined to be circular (223 contigs and 165 MAGs). Around 2022 assemblies (610 contigs and 1412 MAGs) had more than 12 core genes, with 1019 (386 contigs and 657 MAGs) having all 15 core genes. The expanded mitochondrial species tree was built with clusters of metagenomic and reference mitochondrial sequences. In this tree, 140 out of 248 representative mitochondrial genome clusters contained both metagenomic and reference mitochondrial sequences, while 93 clusters contained exclusively metagenomic sequences, which roughly corresponded to 37% of total mitochondrial taxon richness. Clustering of the 3131 assemblies with vclust [48] yielded 881 species-level OTUs. In the phylogenomic tree, most metagenomic mitochondria grouped with Ascomycota and Basidiomycota. 888 mitoMAGs clustered together with fungal mitogenomes of the diverse order Helotiales in the Ascomycota and were found to be present across a wide range of biomes (Fig. 3). Interestingly, we also identified a clade basal to known fungal mitogenomes. Such a group potentially represents a new lineage of fungi or another new unexplored lineage in Opisthokonta.

Mitochondrial gene order changes across genera

To address variation observed in gene order, we looked across a subset of our dataset consisting exclusively of single-contig, circular mitogenome assemblies. By using the *cox1* gene to orient the circular genomes, we applied the Levenshtein distance approach to calculate the pairwise number of shifts in gene order. Overall, we found that gene order remains consistent within genera but changes across genera (Fig. 4).

We investigated five genera in greater detail to explore gene order within a genus: *Cochliobolus* ($n = 6$ strains of *C. heterostrophus*), *Fusarium* ($n = 9$ species), *Suillus* ($n = 11$ species), *Lactarius* ($n = 6$ species), and *Phyllosticta* ($n = 16$ strains from 6 species). While gene order is maintained in four out of five of these genera, in *Phyllosticta* we see a clear separation into two distinct gene order patterns (Supplementary Fig. S7). Strains of *P. citricarpa*, *P. paracitricarpa,* and *P. citriasiana* show one gene order pattern: *cox1-cox2-nad4L-nad5*(with embedded *rps3*)*-nad1-cob-atp6-nad3-nad6-cox3-nad4-nad2-atp9*. Strains of *P. citribraziliensis*, *P. citrichinaensis*, and *P. capitalensis* show another (characterized by the different order and placement of *nad2-nad4*): *cox1-cox2-nad4L-nad5* (with embedded *rps3*)*-nad2-nad4-nad1-cob-atp6-nad3-nad6-cox3-atp9*. These distinct patterns nearly correlate with mitogenome size (itself correlated with lifestyle), except for *P. citricinaensis*, a pathogenic species with a larger mitogenome that shares its gene order with the non-pathogenic species with smaller mitogenomes.

In *C. heterostrophus*, all the assemblies are identical except for six SNPs and three blocks of deletions. This represents ~0.009 snp/kb, compared to ~0.61 snp/kb in the corresponding nuclear genomes. Interestingly, we found that these mutations were correlated with geography. We observed two identical deletions in the Mexican strains PR9b and PR1 × 412, both of which harbor LAGLIDADG endonuclease-like sequences in the other four strains. Such an observation suggests that the HEGs are active, and either their presence in the other four strains is new, or the deletion in these two strains is recent. The first is 1265 bp in an intron of cox3, and the other is 1596 bp in an intron of cob. We also find a 2038 bp deletion in the intergenic region between the *nad2* and *cox2* genes in Hm338, a strain isolated from New York.

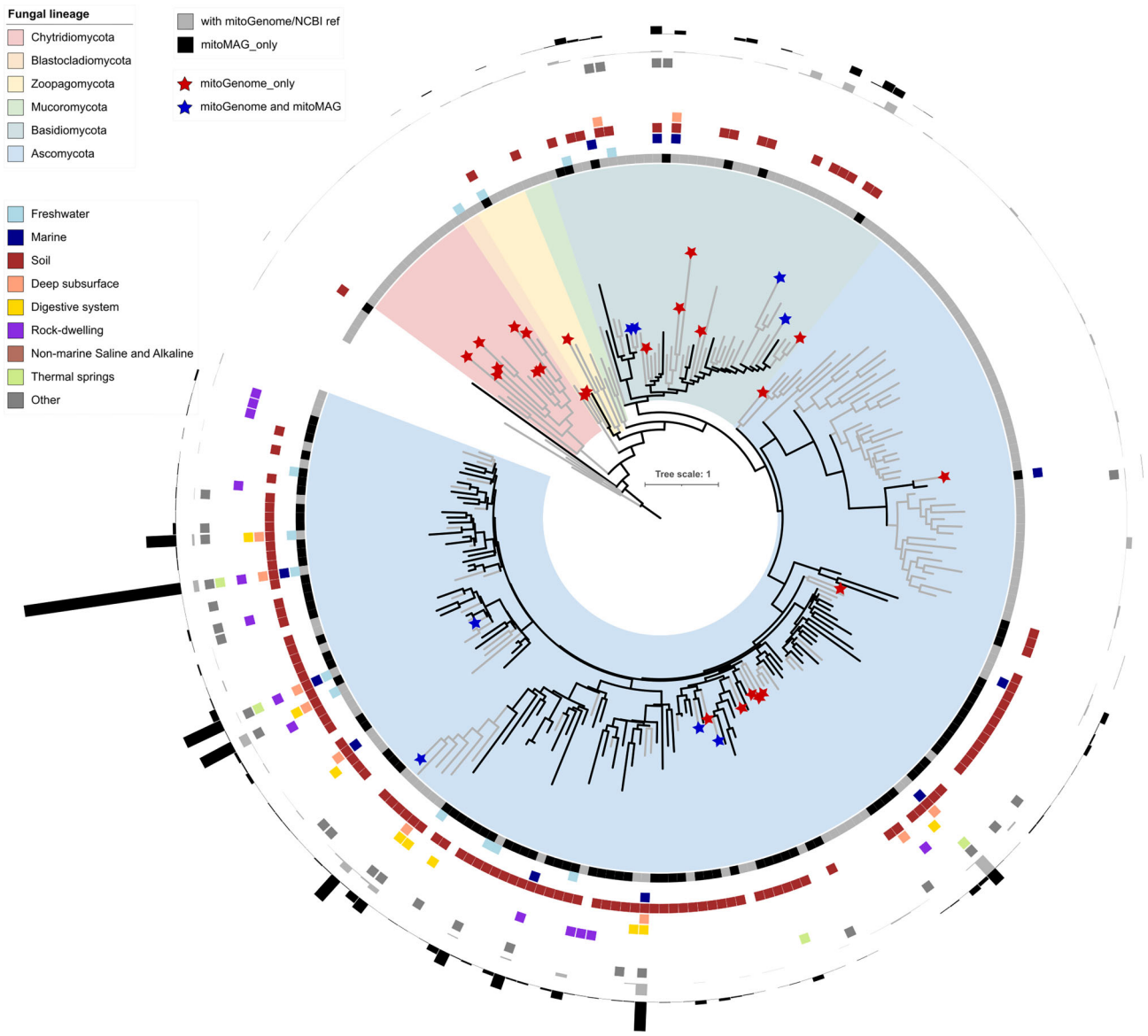


Figure 3. Contribution of mitoMAGs to fungal mitochondrial genome diversity. Phylogenetic tree built from a dereplicated set of mitoMAGs and mitogenomes from other sources. Branches in black indicate clades exclusively containing mitoMAGs. Grey branches contain mitogenomes from NCBI only, NCBI and JGI mitogenome collection (blue stars), or exclusively JGI mitogenomes (red stars). Moving outward from the tree: the innermost ring reproduces this classification. For clades with mitoMAGs, their environmental origin is shown next with separate individual rings. The two outermost rings use bars to indicate the number of genomes in the cluster that are either from mitogenomes (grey) or mitoMAGs (black).

MycoCosm comparative mitogenomic tools

For all mitochondria annotated at the JGI, including those in this study (see https://mycocosm.jgi.doe.gov/fungal_mito_manuscript), we provide mitochondrial genome annotation tabs on individual MycoCosm genome portals (e.g. https://mycocosm.jgi.doe.gov/Lacind1/organelle/mitochondria). Such tabs include overall genome assembly and annotation statistics, as well as a Circos-based [31] visualization showing gene models (colored by type), a line graph of percent GC, and any identified repetitive regions. Additionally, a comparative summary of annotated mitogenomes in MycoCosm is also available, providing a tabular format of assembly statistics (size, number of scaffolds, and GC%), and gene counts (known core genes, hypothetical genes, homing endonucleases, rRNAs, and tRNAs). Such a summary can be found at https://mycocosm.jgi.doe.gov/mycocosm/mito/summary?organism=fungal_mito_manuscript. Finally, while the mitochondrial annotation pipeline currently exists within the larger context of JGI's nuclear genome annotation and analysis capabilities, non-JGI sequenced nuclear and mitochondrial genomes together can be submitted for annotation via https://gold.jgi.doe.gov.

Discussion

Mitochondria are ancient eukaryotic endosymbionts essential for ATP production. The increasing availability of eukaryotic mitochondrial genomes has made informative comparative analyses more routine. Broad analyses in green plants, for example, shed light on repeat size and evolution [59]. In fungi, kingdom-wide analyses remain limited; most often, they remain confined to specific clades. Here, we annotated a combined dataset of nearly 10 000 diverse fungal mitogenomes and pursued in-depth exploration of a high-

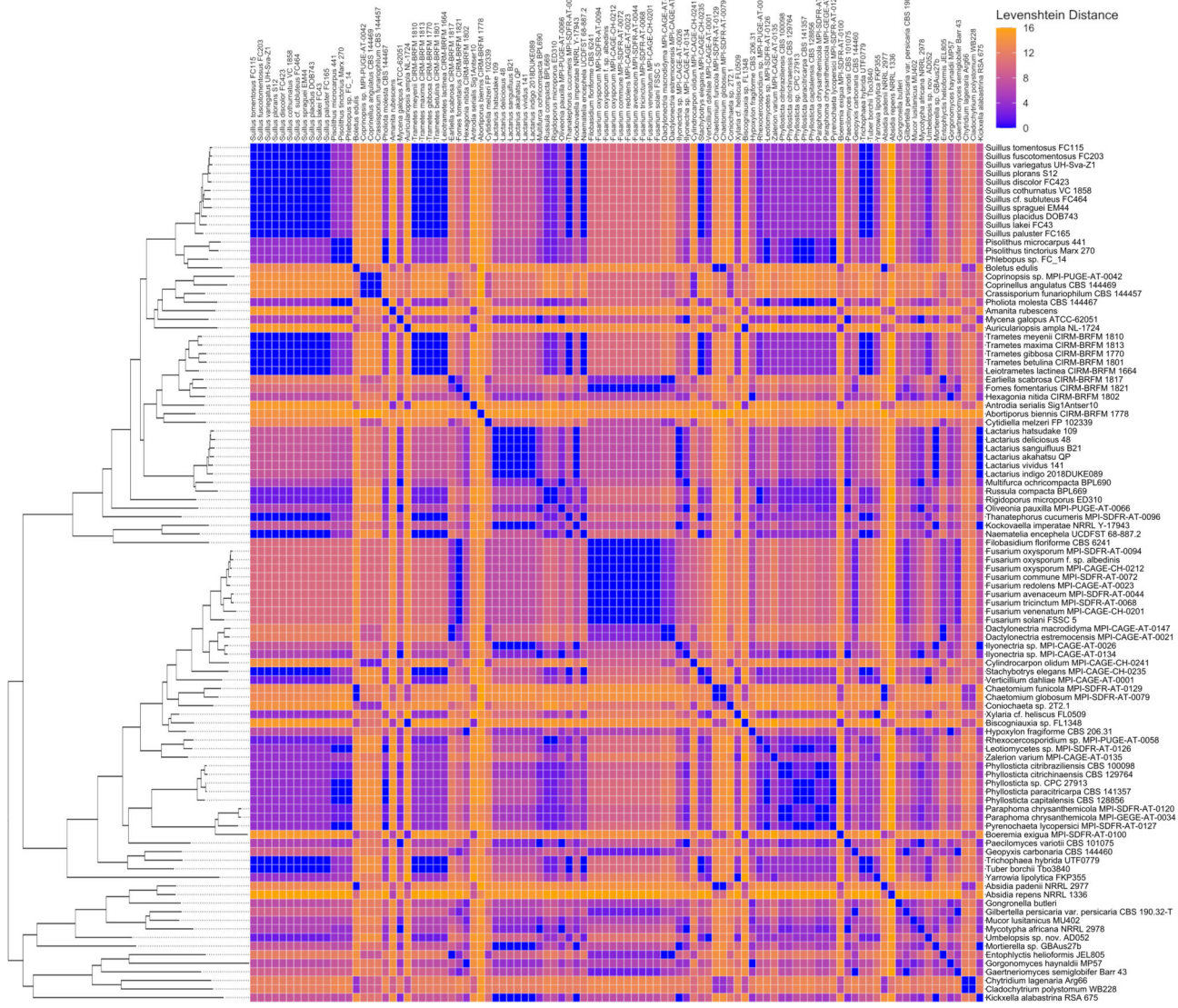


Figure 4. Pairwise Levenshtein distance for 100 circular genomes. Gene order was used to calculate Levenshtein distance for each pair of genomes, reoriented to start at cox1. Gene order is maintained within genera, except in the case of Phyllosticta, which shows two distinct order patterns. Closely related genera do not share gene order, but some distantly related ones do, possibly due to the limited number of permutations leading to convergence.

quality subset of genomes available within the JGI MycoCosm portal. With this information, we detected 15 core protein-coding genes conserved across the kingdom, characterized mitogenome architecture and gene content, and discovered ancient patterns/shifts in mitochondrial tRNA usage as part of this process. We were also able to create a robust mitochondrial phylogeny, which can serve as a backbone for future mitogenome studies. Using this phylogeny, we found substantial agreement between mitochondria and host phylogenies, allowing us to use mitochondria to document the presence of fungi across over 25 000 metagenomic datasets. We also detected 6467 (69.8% of the total 9255) hidden mitochondria within fungal genomes available at the NCBI, revealing a trove of information for future study.

By surveying our phylogenetically diverse dataset, we found that mitogenome architecture is more variable than expected. Most significantly, we observed huge mitogenomes in desert truffles (Pezizales), where sizes were > 1Mb on average. These represent the largest fungal mitogenomes identified to date (the previously reported largest fungal mitogenome, *Morchella crassipes*, being only 531 kb [21]). We also see that the related species, *Morchella importuna*, also from the Pezizales, has a mitogenome size of 274 kb. Taken together, these observations suggest increased plasticity or selective pressure in the Pezizales to maintain large genomes. Such expansion of genome size appears related to increased intron sizes, with additional contribution of more and larger hypothetical proteins. A large proportion of these genomes, however, is intergenic. Size differences were also observed between two subclades of *Phyllosticta*, with notable correlation with pathogenicity, where longer mitogenomes were observed in pathogenic strains. Similar interspecies variations in mitogenome sizes have been reported in other phytopathogenic fungi as well, such as species of *Monilinia* which infect stone and pome fruits [60].

We see a dynamic range of GC content across mitogenomes that did not necessarily follow nuclear GC content trends in their hosts at the subphylum level. We also found several

clades with higher average GC% than expected, for example, Chytridiomycota, Mucoromycotina, and Pucciniomycotina. Thus, while low GC content thresholds have been suggested as useful for identifying mitochondrial scaffolds [12], our overall assessment of GC% trends suggests that these standards may not be consistently applicable across the kingdom.

Using a high-quality set of 300 + mostly single-scaffold fungal mitochondrial genomes, we find general conservation of the 14 genes associated with oxidative phosphorylation complexes I, II, IV, and V across the kingdom. Such a finding is consistent with other analyses for individual fungal clades [61]. We observed that the order of genes changes across genera, consistent with a previous study [62] reporting variability between and within the major phyla, but which made no comment on conservation at the genus level. Other work, for example, in vertebrates, has shown gene order consistency at the order level [63]. Our finding that gene order is conserved within fungal genera can add to the broader discussion of mitochondrial gene order variability in fungi. In particular, the fact that we observe discrete boundaries between fungal genera with a mitochondrial phylogeny supports the taxonomic validity observed with nuclear phylogenomic analyses.

We also detected 1574 and 2022 mitogenomes from NCBI and IMG metagenomes, respectively, with ≥ 12 core genes. In general, we found that subunits of cytochrome c oxidase (Complex IV) and cytochrome bc1 (Complex III) are the most conserved across fungi, while rps3 is the most variable. Subunits of the NADH:ubiquinone oxidoreductase (Complex I) appear missing from Cryptomycota, despite being otherwise generally conserved. Similarly, ATP synthase (Complex V) subunits atp8 and atp9 are missing from mitogenomes of most Dothideomycetes, with evidence suggesting that while atp8 may have been transferred to the nucleus in only Dothideomycetes, atp9 might represent potential transfer in progress for a broader collection of groups. Our data shows that the presence of atp9 in the nuclear genomes is primarily restricted to the Pezizomycotina but is not found in the Pezizales, which is the earliest branching order in the Pezizomycotina. The presence of the nuclear atp9 seems to have allowed specific losses of mitochondrial atp9 in some Pezizomycotina genomes, such as in some members of the Aulographales and some *Chaetomium* species. Previous work has shown that atp8 is missing from the mitochondrial genomes of non-fungal eukaryotes [64–66], as well as only smaller groups of fungi [51–[53]. Due to the small size and divergent evolution of atp8, others have argued that the gene may be simply missed during annotation [67].

The atp9 component of the ATP synthase-F0 is not essential for survival in some yeasts [68], and it can be experimentally transferred to the nucleus [52]. The experimental relocation required multiple steps, including changes to the coding sequence, targeting signals, and transmembrane domains. Our observed pattern, where atp9 is present in both the mitochondrial and nuclear genomes in some clades, may represent transitional states where the gene is actively being moved to the nucleus, but full functionality has not been achieved yet, which would eventually allow the deletion of the mitochondrial copy. We find that while rps3 is absent from mitogenomes of both Chytridiomycota and Mucoromycota lineages, a nuclear copy is found only in the latter. While rare, we find other examples of nuclear-encoded mitochondrial genes, some of which have supporting mitochondrial targeting signals. However, none of these organisms show corresponding losses of these genes from the mitochondrial genomes, and, given that our dataset lacks close relatives, we can speculate that these may represent transitions in progress, but additional sequencing might be necessary to establish any broader evolutionary patterns.

We also observed a significant phylogenetic signal of cospeciation ($P < 0.01$) between mitochondria and their hosts, in agreement with findings in other eukaryotes despite the fact that fungi engage in a variety of non-canonical interactions where hyphae from one or more partners can fuse and exchange genetic material. Combining this observation with our new detection and annotation tools, we anticipate that mining of resequencing data and samples with degraded DNAs (e.g. herbarium specimens) could provide a major source of new mitochondrial information for exploring biology and taxonomic relationships amongst fungi. In this analysis, we recovered over 3000 fungal mitogenomes from metagenome datasets across a diversity of environmental habitats. We observed that the majority of metagenomic mitochondria clustered with the Helotiales in the Ascomycota, consistent with certain studies suggesting that Ascomycota species are most abundant across multiple environmental types [69]. However, we also see representation of a clade basal to known fungal mitogenomes. Given that new Opisthokont clades continue to be described [70], such results are promising targets for applying our approach for further in-depth exploration of additional phylogenetic novelty. Given the limitations of using ITS as a fungal marker sequence [71], these results demonstrate that mitochondrial genes might offer an attractive alternative set of markers for microbial community-level analyses, even if support values might be relatively lower. Indeed, such approaches have been implemented previously in soil-inhabiting invertebrates [72], however, no large-scale approach has yet been reported for Fungi. By augmenting or replacing existing nuclear markers with those derived from mitogenomes, researchers can sift existing datasets to identify previously unsampled fungal lineages, or potentially explore interactions to identify the abundance of eukaryotes in connection with bacteria.

Related to the development of marker genes, while such identity can be made on a genus level, the SNP / indel results seen in *Cochliobolus heterostrophus* suggest that differences exist between strains and appear to be related to geographic origin. Similarly, we see SNP and indel differences between strains of the same species in the *Phyllosticta* study group. Indeed, current annotation data from geographically distinct strains across the fungal kingdom remains sparse. However, as such resources grow with additional sampling, it may become feasible to use mitogenes as markers to track the movement of species and even strains within metagenome datasets, as well as provide additional support for geographic origins.

The patterns of mitochondrially encoded tRNA genes within Fungi have previously been explored, but in a limited context, for example: general reductions in a few species of Chytridiomycota and differences in model organisms such as *Neurospora crassa*. Thus, it remains a poorly explored area of fungal mitochondria regulation, particularly when examined in the context of codon usage patterns. We analyzed tRNA patterns across the kingdom and found a general encoding of a single tRNA per amino acid, consistent with the utilization

of wobble-pairing to recruit amino acids during translation and the import of additional tRNAs from their hosts [73]. Consistent with the previous limited observations in *Spizellomyces punctatus* [58], *Hyaloraphidium curvatum* [74], and four monoblepharidaleans [75], we find evidence suggesting that the entire Chytridiomycota phylum contains a highly reduced set of tRNAs, with additional further reduction in the order Rhizophydiales. These absences suggest a specific loss event during the emergence of the Chytridiomycota, as early-diverging non-fungal eukaryotes exhibit a similar pattern to non-chytrid Fungi with a few exceptions. Interestingly, included in the set of tRNAs is a unique suppressor, corresponding to the use of the Chlorophycean Mitochondrial Code (code 16) in some members of the phylum, an intriguing observation given that many chytrids are also frequent pathogens of chlorophytes [76]. Continuing through the tree, we found the replacement of $tRNA^{Arg}$(UCU) with $tRNA^{Arg}$(ACG) as well as two losses of $tRNA^{Trp}$ corresponding to the Russulales in the Basidiomycota, as well as the emergence of the Ascomycota. These suggest major shifts in mitochondrial translation regulation. We found several cases where predicted tRNAs exist, but no corresponding codons could be detected, which was most frequently observed for $tRNA^{Cys}$(GCA). Given that nonessential genes are often rapidly pseudogenized and lost in obligate endosymbiotic systems [77], the retention of such tRNAs across evolutionary time warrants further exploration. While shifts in nuclear codon usage have been documented previously [78, 79], our work can help assess fungal mitochondrial tRNA shifts and whether they follow patterns shown for nuclear tRNAs.

The resources and utilities made available in this work should provide opportunities for the fungal and/or mitochondrial research community, irrespective of level of familiarity with JGI's MycoCosm platform. Similarly, the analyses presented here provide a solid foundation for growth in a variety of research directions. Our results highlight specific fungal clades that demand further exploration of various biological mechanisms, such as coding potential. More broadly, the increasing ease of mitogenomic sequencing and observed kingdom-wide conservation of marker genes, coupled with phylogenetic congruence of nuclear and mitochondrial genome phylogeny, together provide an attractive path toward the future of fungal metagenomics.

Acknowledgements

We would like to thank the following collaborators for permission to incorporate their unpublished genome data into our analyses: Francis Martin and Gabor Kovacs for *Mattirolomyces terfezioides*, and Allison Walker for *Peniophora* sp. CONTA. Despite repeated attempts to contact Spencer Stong prior to submission, we were unable to reach him. As a result, Spencer Stong did not have the opportunity to review or approve the final accepted version and, therefore, cannot be held responsible for the content post-publication.

Author contributions: S.R.A.: Writing, Investigation. S.H.: Writing, Investigation, Methodology. S.S.: Methodology. A.Sa.: Software, Methodology, Investigation. A.St.: Writing, Investigation. K.L.: Methodology. R.R.: Methodology. I.S.: Software, Data curation. I.L.: Software, Data curation. S.D.: Software, Data curation. F.S.: Writing, Investigation. M.F.R.: Investigation. J.C.V.: Investigation. I.V.G.: Editing, Supervision. S.J.M.: Writing, Investigation, Supervision.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

The work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

Data availability

The JGI-derived mitogenome assemblies and annotations are available at https://mycocosm.jgi.doe.gov and can be accessed using the genome portal IDs (and associated individual URLs) listed in Supplementary Table S2. The NCBI fungal nuclear genome assemblies were accessed from https://www.ncbi.nlm.nih.gov using the accessions provided in Supplementary Table S3. The IMG metagenome assemblies were accessed from https://img.jgi.doe.gov using the accessions provided in Supplementary Table S4. The NCBI-derived and IMG-derived mitogenome assemblies and annotations can be accessed from https://mycocosm.jgi.doe.gov/fungal_mito_manuscript.

References

1. Karnkowska A, Vacek V, Zubáčová Z *et al.* A eukaryote without a mitochondrial organelle. *Curr Biol* 2016;26:1274–84. https://doi.org/10.1016/j.cub.2016.03.053
2. Friedman JR, Nunnari J Mitochondrial form and function. *Nature* 2014;505:335–43. https://doi.org/10.1038/nature12985
3. Sloan DB, Warren JM, Williams AM *et al.* Cytonuclear integration and co-evolution. *Nat Rev Genet* 2018;19:635–48. https://doi.org/10.1038/s41576-018-0035-9
4. Nunnari J, Suomalainen A Mitochondria: in sickness and in health. *Cell* 2012;148:1145–59. https://doi.org/10.1016/j.cell.2012.02.035
5. Tobler M, Barts N, Greenway R Mitochondria and the origin of species: bridging genetic and ecological perspectives on speciation processes. *Integr Comp Biol* 2019;59:900–11. https://doi.org/10.1093/icb/icz025
6. Megarioti AH, Kouvelis VN The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol Evol* 2020;12:1337–54. https://doi.org/10.1093/gbe/evaa126
7. Knight RD, Landweber LF, Yarus M How mitochondria redefine the code. *J Mol Evol* 2001;53:299–313. https://doi.org/10.1007/s002390010220
8. Lang BF, Beck N, Prince S *et al.* 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
9. Bernt M, Donath A, Jühling F *et al.* MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* 2013;69:313–9. https://doi.org/10.1016/j.ympev.2012.08.023
10. Meng G, Li Y, Yang C *et al.* MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res* 2019;47:e63. https://doi.org/10.1093/nar/gkz173

11. Wyman SK, Jansen RK, Boore JL Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2004;20:3252–5. https://doi.org/10.1093/bioinformatics/bth352
12. Roger AJ, Muñoz-Gómez SA, Kamikawa R The origin and diversification of mitochondria. *Curr Biol* 2017;27:R1177–92. https://doi.org/10.1016/j.cub.2017.09.015
13. Keeling PJ, Campo JD Marine protists are not just big bacteria. *Curr Biol* 2017;27:R541–9. https://doi.org/10.1016/j.cub.2017.03.075
14. Alexander H, Hu SK, Krinos AI *et al.* Eukaryotic genomes from a global metagenomic dataset illuminate trophic modes and biogeography of ocean plankton. *mBio* 2023;14:e01676–23. https://doi.org/10.1128/mbio.01676-23
15. Freel KC, Friedrich A, Schacherer J Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res* 2015;15:fov023. https://doi.org/10.1093/femsyr/fov023
16. Hugaboom M, Hatmaker EA, LaBella AL *et al.* Evolution and codon usage bias of mitochondrial and nuclear genomes in Aspergillus section Flavi. *G3* 2023;13:jkac285. https://doi.org/10.1093/g3journal/jkac285
17. Fonseca PLC, De-Paula RB, Araújo DS *et al.* Global characterization of fungal mitogenomes: new insights on genomic diversity and dynamism of coding genes and accessory elements. *Front Microbiol* 2021;12:787283.
18. Barr CM, Neiman M, Taylor DR Inheritance and recombination of mitochondrial genomes in plants, fungi and animals: research review. *New Phytol* 2005;168:39–50. https://doi.org/10.1111/j.1469-8137.2005.01492.x
19. Haridas S, Salamov A, Grigoriev IV. Fungal Genome Annotation. In: de Vries R. P., Tsang A., Grigoriev I. V. (eds), *Fungal Genomics: Methods and Protocols*. New York, NY: Springer 2018, pp.171–84.
20. Grigoriev IV, Nikitin R, Haridas S *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* 2014;42:D699–704. https://doi.org/10.1093/nar/gkt1183
21. Liu W, Cai Y, Zhang Q *et al.* Subchromosome-scale nuclear and complete mitochondrial genome characteristics of morchella crassipes. *Int J Mol Sci* 2020;21:483.
22. Chen I-MA, Chu K, Palaniappan K *et al.* The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res* 2023;51:D723–32. https://doi.org/10.1093/nar/gkac976
23. Zerbino DR, Birney E Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;18:821–9. https://doi.org/10.1101/gr.074492.107
24. Li H, Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324
25. Gnerre S, Maccallum I, Przybylski D *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 2010;108:1513–8. https://doi.org/10.1073/pnas.1017351108
26. Lin Y, Yuan J, Kolmogorov M *et al.* Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci USA* 2016;113:E8396–405. https://doi.org/10.1073/pnas.1604560113
27. Eddy SR Multiple alignment using hidden markov models. *Proc Int Conf Intell Syst Mol Biol* 1995;3:114–20.
28. Paquin B, Laforest M-J, Forget L, Longcore J *et al.* The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr Genet* 1997;31:380–95.
29. Lowe TM, Eddy SR tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997;25:955–64. https://doi.org/10.1093/nar/25.5.955
30. Nawrocki EP, Burge SW, Bateman A *et al.* Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* 2015;43:D130–7. https://doi.org/10.1093/nar/gku1063
31. Krzywinski M, Schein J, Birol I *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639–45. https://doi.org/10.1101/gr.092759.109
32. Edgar RC, Drive RM, Valley M MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–7. https://doi.org/10.1093/nar/gkh340
33. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–3. https://doi.org/10.1093/bioinformatics/btp348
34. Minh BQ, Schmidt HA, Chernomor O *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015
35. Mistry J, Chuguransky S, Williams L *et al.* Pfam: the protein families database in 2021. *Nucleic Acids Res* 2021;49:D412–9. https://doi.org/10.1093/nar/gkaa913
36. Van Dongen S Graph clustering via a discrete uncoupling process. *SIAM J Matrix Anal Appl* 2008;30:121–41. https://doi.org/10.1137/040608635
37. Steinegger M, Söding J MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;35:1026–8. https://doi.org/10.1038/nbt.3988
38. Katoh K, Standley DM MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–80. https://doi.org/10.1093/molbev/mst010
39. Castresana J Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000;17:540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334
40. Price MN, Dehal PS, Arkin AP FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;26:1641–50. https://doi.org/10.1093/molbev/msp077
41. Nawrocki EP, Eddy SR Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 2013;29:2933–5. https://doi.org/10.1093/bioinformatics/btt509
42. Kalvari I, Nawrocki EP, Ontiveros-Palacios N *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res* 2021;49:D192–200. https://doi.org/10.1093/nar/gkaa1047
43. Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acid Res* 2013;41:D590–6. https://doi.org/10.1093/nar/gks1219
44. Guillou L, Bachar D, Audic S *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 2013;41:D597–604. https://doi.org/10.1093/nar/gks1160
45. Kang DD, Li F, Kirton E *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;7:e7359. https://doi.org/10.7717/peerj.7359
46. Muñoz-Gómez SA, Susko E, Williamson K *et al.* Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat Ecol Evol* 2022;6:253–62
47. Price MN, Dehal PS, Arkin AP FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490. https://doi.org/10.1371/journal.pone.0009490
48. Rognes T, Flouri T, Nichols B *et al.* VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016;4:e2584. https://doi.org/10.7717/peerj.2584
49. James TY, Pelin A, Bonen L *et al.* Shared signatures of parasitism and phylogenomics unite Cryptomycota and Microsporidia. *Curr Biol* 2013;23:1548–53. https://doi.org/10.1016/j.cub.2013.06.057
50. Haag KL, James TY, Pombert J-F *et al.* Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proc Natl Acad Sci USA* 2014;111:15480–5. https://doi.org/10.1073/pnas.1410442111

51. Déquard-Chablat M, Sellem CH, Golik P *et al.* Two nuclear life cycle–regulated genes encode interchangeable subunits c of mitochondrial ATP synthase in Podospora anserina. *Mol Biol Evol* 2011;28:2063–75. https://doi.org/10.1093/molbev/msr025
52. Bietenhader M, Martos A, Tetaud E *et al.* Experimental relocation of the mitochondrial ATP9 gene to the nucleus reveals forces underlying mitochondrial genome evolution. *PLoS Genet* 2012;8:e1002876. https://doi.org/10.1371/journal.pgen.1002876
53. Boominathan A, Vanhoozer S, Basisty N *et al.* Stable nuclear expression of ATP8 and ATP6 genes rescues a mtDNA Complex V null mutant. *Nucleic Acids Res* 2016;44, pp.9342–57.
54. Korovesi AG, Ntertilis M, Kouvelis VN Mt-rps3 is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. *Mol Phylogenet Evol* 2018;127:74–86. https://doi.org/10.1016/j.ympev.2018.04.037
55. Thumuluri V, Almagro Armenteros JJ, Johansen AR *et al.* DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res* 2022;50:W228–34. https://doi.org/10.1093/nar/gkac278
56. Allio R, Donega S, Galtier N *et al.* Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol* 2017;34:2762–72. https://doi.org/10.1093/molbev/msx197
57. Latorre A, Manzano-Marín A Dissecting genome reduction and trait loss in insect endosymbionts. *Ann NY Acad Sci* 2017;1389:52–75. https://doi.org/10.1111/nyas.13222
58. Laforest MJ, Roewer I, Lang BF Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and UAG 'stop' codons recognized as leucine. *Nucleic Acids Res* 1997;25:626–32. https://doi.org/10.1093/nar/25.3.626
59. Wynn EL, Christensen AC Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3* 2019;9:549–59. https://doi.org/10.1534/g3.118.200948
60. Yildiz G, Ozkilinc H Pan-mitogenomics approach discovers diversity and dynamism in the prominent brown rot fungal pathogens. *Front Microbiol* 2021;12:647989. https://doi.org/10.3389/fmicb.2021.647989
61. Fonseca PLC, Badotti F, De-Paula RB *et al.* Exploring the relationship among divergence time and coding and non-coding elements in the shaping of fungal mitochondrial genomes. *Front Microbiol* 2020;11:765. https://doi.org/10.3389/fmicb.2020.00765
62. Aguileta G, de Vienne DM, Ross ON *et al.* High variability of mitochondrial gene order among fungi. *Genome Biol Evol* 2014;6:451–65. https://doi.org/10.1093/gbe/evu028
63. Montaña-Lozano P, Moreno-Carmona M, Ochoa-Capera M *et al.* Comparative genomic analysis of vertebrate mitochondrial reveals a differential of rearrangements rate between taxonomic class. *Sci Rep* 2022;12:5479. https://doi.org/10.1038/s41598-022-09512-2
64. Solà E, Álvarez-Presas M, Frías-López C *et al.* Evolutionary analysis of mitogenomes from parasitic and free-living flatworms. *PLoS One* 2015;10:e0120081. https://doi.org/10.1371/journal.pone.0120081
65. Kim T, Kern E, Park C *et al.* The bipartite mitochondrial genome of ruizia karukerae (Rhigonematomorpha, Nematoda). *Sci Rep* 2018;8:7482. https://doi.org/10.1038/s41598-018-25759-0
66. Shimada D, Hiruta SF, Takahoshi K *et al.* Does atp8 exist in the mitochondrial genome of proseriata (Metazoa: platyhelminthes)? *Animal Genet* 2023;30:200161. https://doi.org/10.1016/j.angen.2023.200161
67. Breton S, Stewart DT, Hoeh WR Characterization of a mitochondrial ORF from the gender-associated mtDNAs of Mytilus spp. (Bivalvia: mytilidae): identification of the "missing" ATPase 8 gene. *Mar Geonomics* 2010;3:11–8. https://doi.org/10.1016/j.margen.2010.01.001
68. Clark-Walker GD, Francois F, Chen XJ *et al.* Mitochondrial ATP synthase subunit 9 is not required for viability of the petite-negative yeast *Kluyveromyces Lactis*. *Curr Genet* 1997;31:488–93. https://doi.org/10.1007/s002940050234
69. de Mesquita CPB, Vimercati L, Wu D *et al.* Fungal diversity and function in metagenomes sequenced from extreme environments. *Fungal Ecol* 2024;72:101383. https://doi.org/10.1016/j.funeco.2024.101383
70. Arroyo AS, López-Escardó D, Kim E *et al.* Novel diversity of deeply branching holomycota and unicellular holozoans revealed by metabarcoding in middle Paraná river. *Argentina Front Ecol Evol* 2018;6:378849. 10.3389/fevo.2018.00099
71. Kauserud H ITS alchemy: on the use of ITS as a DNA marker in fungal ecology. *Fungal Ecol* 2023;65:101274. https://doi.org/10.1016/j.funeco.2023.101274
72. Andújar C, Arribas P, Ruzicka F *et al.* Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular ecology* 2015;24:3603–17.
73. Schneider A Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annu Rev Biochem* 2011;80:1033–53. https://doi.org/10.1146/annurev-biochem-060109-092838
74. Forget L, Ustinova J, Wang Z *et al.* Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Mol Biol Evol* 2002;19:310–9. https://doi.org/10.1093/oxfordjournals.molbev.a004084
75. Bullerwell CE, Forget L, Lang BF Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. *Nucleic Acids Res* 2003;31:1614–23. https://doi.org/10.1093/nar/gkg264
76. Carney LT, Lane TW Parasites in algae mass culture. *Front Microbiol* 2014;5:278. https://doi.org/10.3389/fmicb.2014.00278
77. Wixon J Reductive evolution in Bacteria: buchnera sp., Rickettsia prowazekii and Mycobacterium leprae. *Int J Genomics Proteomics* 2001;2:44–8.
78. Shen XX, Steenwyk JL, LaBella AL *et al.* Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum. *Sci Adv* 2020;6:eabd0079. https://doi.org/10.1126/sciadv.abd0079
79. Labella AL, Opulente DA, Steenwyk JL *et al.* Variation and selection on codon usage bias across an entire subphylum. *PLoS Genet* 2019;15, p.e1008304. https://doi.org/10.1371/journal.pgen.1008304

Received: May 20, 2024. **Revised:** October 15, 2025. **Accepted:** November 14, 2025
Published by Oxford University Press 2026.