Check for updates

Mitochondrial genome diversity across the subphylum Saccharomycotina

John F. Wolters[1], Abigail L. LaBella[2,3], Dana A. Opulente[1,4], Antonis Rokas[3] and Chris Todd Hittinger[1]*

[1]Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI, United States, [2]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, United States, [3]Department of Biological Sciences, Vanderbilt University, Nashville, TN, United States, [4]Biology Department, Villanova University, Villanova, PA, United States

Introduction: Eukaryotic life depends on the functional elements encoded by both the nuclear genome and organellar genomes, such as those contained within the mitochondria. The content, size, and structure of the mitochondrial genome varies across organisms with potentially large implications for phenotypic variance and resulting evolutionary trajectories. Among yeasts in the subphylum Saccharomycotina, extensive differences have been observed in various species relative to the model yeast *Saccharomyces cerevisiae*, but mitochondrial genome sampling across many groups has been scarce, even as hundreds of nuclear genomes have become available.

Methods: By extracting mitochondrial assemblies from existing short-read genome sequence datasets, we have greatly expanded both the number of available genomes and the coverage across sparsely sampled clades.

Results: Comparison of 353 yeast mitochondrial genomes revealed that, while size and GC content were fairly consistent across species, those in the genera *Metschnikowia* and *Saccharomyces* trended larger, while several species in the order Saccharomycetales, which includes *S. cerevisiae*, exhibited lower GC content. Extreme examples for both size and GC content were scattered throughout the subphylum. All mitochondrial genomes shared a core set of protein-coding genes for Complexes III, IV, and V, but they varied in the presence or absence of mitochondrially-encoded canonical Complex I genes. We traced the loss of Complex I genes to a major event in the ancestor of the orders Saccharomycetales and Saccharomycodales, but we also observed several independent losses in the orders Phaffomycetales, Pichiales, and Dipodascales. In contrast to prior hypotheses based on smaller-scale datasets, comparison of evolutionary rates in protein-coding genes showed no bias towards elevated rates among aerobically fermenting (Crabtree/Warburg-positive) yeasts. Mitochondrial introns were widely distributed, but they were highly enriched in some groups. The majority of mitochondrial introns were poorly conserved within groups, but several were shared within groups, between groups, and even across taxonomic orders, which is consistent with horizontal gene transfer, likely involving homing endonucleases acting as selfish elements.

Discussion: As the number of available fungal nuclear genomes continues to expand, the methods described here to retrieve mitochondrial genome sequences from these datasets will prove invaluable to ensuring that studies of fungal mitochondrial genomes keep pace with their nuclear counterparts.

KEYWORDS

yeast, mitochondria, evolution, selection, diversity

1 Introduction

Eukaryotic evolution is a history of multiple genomes coming together. The acquisition of the mitochondria via endosymbiosis enabled new metabolic capacities, but it required the coevolution of two distinct genomes over time and created a novel dynamic (Muñoz-Gómez et al., 2015; Zachar and Szathmáry, 2017). In the vast majority of eukaryotic organisms, the mitochondrial genome (mtDNA) has been vastly reduced to encode a small number of respiratory proteins and their corresponding translational machinery (Johnston and Williams, 2016). All other ancestral mitochondrial genes were either lost or transferred to the nuclear genome, which encodes nearly all genes required for the various mitochondrial functions (Adams and Palmer, 2003). Among extant mtDNAs, there is considerable variation in specific gene content, genome structure, and idiosyncrasies of gene expression (Santamaria et al., 2007; Gualberto et al., 2014; Hao, 2022; Dowling and Wolff, 2023). Dense sampling of eukaryotic taxa is required to understand how this variation arises and its impacts on the evolution and function of both genomes.

Budding yeasts of the subphylum Saccharomycotina (hereafter, yeasts) provide a valuable model for exploring this variation further. The early sequencing of the mtDNA of the model yeast *Saccharomyces cerevisiae* provided a contrast to the picture of mitochondrial evolution that was emerging from animal studies. Whereas most animal mtDNAs were found to be highly gene-dense, small at typically under 20 kb (Santamaria et al., 2007), and lacking introns, the *S. cerevisiae* mtDNA was several times larger (~75–85 kb), contained fewer genes due to lacking any of the canonical mitochondrially-encoded components of Complex I of the electron transport chain, and contained introns in several genes (Foury et al., 1998). Further studies of other eukaryotic groups confirmed that marked differences from the smaller genome seen in animals are the norm (Gualberto and Newton, 2017; Roger et al., 2017; Sandor et al., 2018). The addition of mtDNAs from other yeasts showed that differences in genome size were widespread and that many yeast mtDNAs still encoded a canonical Complex I (Freel et al., 2015; Xiao et al., 2017). However, the current sampling of yeast mtDNAs (Christinaki et al., 2022) remains heavily tilted towards yeasts in the order Saccharomycetales, which contains *S. cerevisiae*, and the order Serinales, which contains the opportunistic pathogen *Candida albicans* (Butler et al., 2009), but these are only two of the 12 orders in the 400-million-year-old subphylum Saccharomycotina (Shen et al., 2018; Groenewald et al., 2023).

Yeasts have become an important model for studying the dynamics of genome evolution and, in particular, its interplay with metabolism (Scannell et al., 2011; Hittinger, 2013; Hittinger et al., 2015; Opulente et al., 2018). Nuclear genome sequences for hundreds of species across all major clades within Saccharomycotina are now available (Shen et al., 2018). However, the availability of mtDNAs for this subphylum is comparatively lacking. In this work, we demonstrate that yeast mtDNAs can be recovered from publicly available short-read genome sequencing datasets, and we more than doubled the number of available mitochondrial genomes across the subphylum to 353 mtDNAs. We show that there is considerable variation in genome size, GC content, patterns of selection, and intron content. We expand the list of species that have independently lost Complex I (Schikora-Tamarit et al., 2021) by reporting several additional cases. Comparisons of gene content revealed that, while there was a major loss of Complex I in the evolution of the ancestor of the orders Saccharomycetales and Saccharomycodales, there are several additional independent losses in other orders. This dataset provides new opportunities to better understand mitochondrial evolution and its relationship to nuclear genome evolution.

2 Results

2.1 Mitochondrial genome sequence mining

To expand the availability of mtDNAs across the subphylum Saccharomycotina, we used a two-pronged approach: first searching for mitochondrial sequences in existing genome assemblies, followed by constructing new genome assemblies using assemblers specialized in generating organellar genomes from short sequencing reads. By searching for matches to existing references, we identified a treasure trove of mitochondrial sequences within the existing assemblies with sizes in the expected ranges for mtDNAs and with elevated coverage relative to the rest of the assembly, which would be consistent with the high copy number expected for the mtDNA (Solieri, 2010; Figure 1). The success rate for extracting nearly complete mtDNAs was quite high for newer assemblies, but it was lower for older assemblies due to either lack of coverage, previously applied computational filters to remove mtDNA, or potentially the use of strains lacking mtDNA to reduce sequencing costs (Supplementary Figure 1). When raw DNA sequencing reads were readily available, reassembly by targeting mitochondrial sequences proved to be even more effective. Out of 232 species assessed via both approaches, 19 were best assembled within the nuclear assembly, whereas 212 were best completed via reassembly [38 by plasmidSPAdes (Antipov et al., 2016) and 174 by NOVOPlasty (Dierckxsens et al., 2017)]. After reducing the mitochondrial genome assemblies to the best representative for each species (Supplementary Table 1), the number of Saccharomycotina species with mtDNAs available increased from 132 (Christinaki et al., 2022) to 353, which included dramatically improved representation in several clades (Figure 2). Many of these mtDNAs were assembled as a circle, but a small number of assemblies remained fragmented, which resulted in missing portions with contig breakpoints that occasionally overlapped annotated genes. The pipeline for searching existing genome assemblies for mitochondrial sequences is available here: https://github.com/JFWolters/IdentifyMitoContigs.

2.2 Phylogeny and genome characteristics

We constructed a phylogeny of yeast mtDNAs based on concatenation of the core protein-coding genes (Figure 3). Overall concordance with the existing nuclear phylogeny was reasonably high (normalized Robinson-Foulds distance 0.24 between matched subtrees). Placement of the recently described (Groenewald et al., 2023) taxonomic orders [previously designated as major clades (Shen et al., 2018)] was consistent between the phylogenies, barring three exceptions: two *Trigonopsis* species grouped closer to Lipomycetales than other Trigonopsidales; the Alaninales were paraphyletic with respect to the Pichiales, rather than forming a single monophyletic outgroup; and the placement of the faster-evolving lineage of
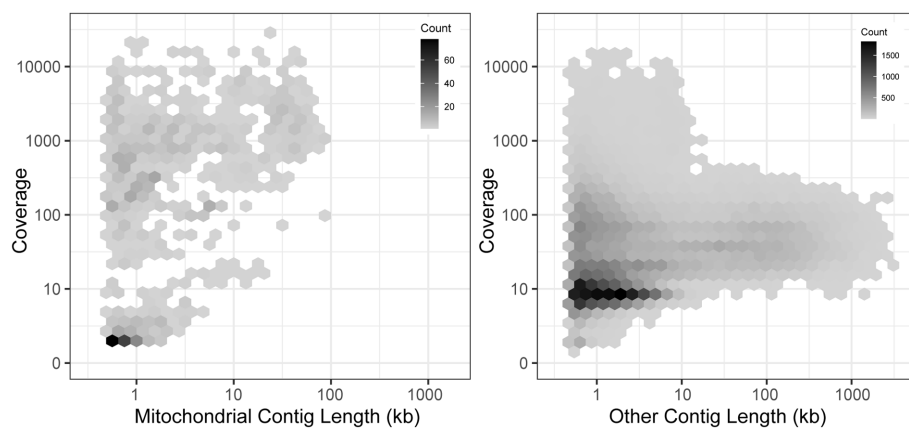
FIGURE 1
Mitochondrial contig profile. The coverage and length profile of contigs from 196 assemblies newly sequenced in (Shen et al., 2018) that were flagged as putative mitochondrial contigs versus all other contigs is displayed ($\log_{10}$ scaling). The most useful mitochondrial contigs generally have a profile of elevated coverage with sizes between 10 and 100 kb, a combination rarely found in other contigs, although strict diagnostic cutoffs are not evident. Many poor-quality putative mitochondrial contigs were found in nuclear genome assemblies, but these were not present in mitochondrially-focused reassemblies.
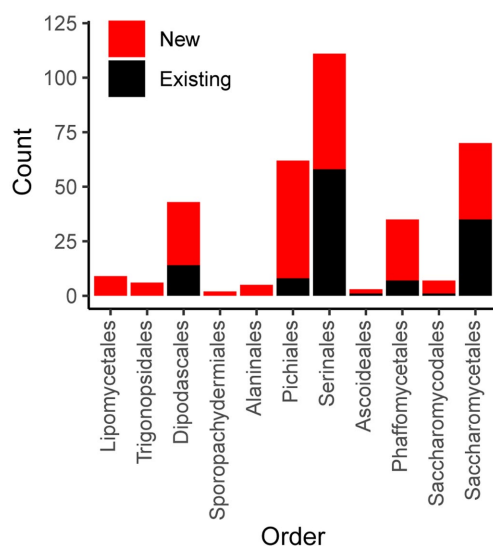

FIGURE 2
Mitochondrial genome counts by taxonomic order. The count of genomes for both newly added and existing genomes from public repositories are displayed according to taxonomic order (classifications recently described by Groenewald et al., 2023). For nearly all orders, a majority of genomes are new [barring Saccharomycetales (35 new versus 35 existing) and Serinales (53 new versus 58 existing)].

Hanseniaspora (order Saccharomycodales) was uncertain due to the long branch at the root of this order. A similar inconsistency was observed in prior phylogenetic analysis where *Hanseniaspora* mtDNAs clustered with the order Serinales (Christinaki et al., 2022). The uncertainty in the placement of the faster-evolving lineage of *Hanseniaspora* is likely due to long branch attraction (Bergsten, 2005). Thus, in Figure 3, we have displayed results from a tree-building run that recovered the order Saccharomycodales as monophyletic, as expected from the genome-scale nuclear phylogeny (Shen et al., 2018).

Within taxonomic orders, groupings of genera were highly congruent with the genome-wide species phylogeny, but some inconsistencies remained in the placements of genera. For example, *Eremothecium* mtDNAs appeared as an outgroup to other Saccharomycetales, rather than grouping with *Kluyveromyces* and *Lachancea* as expected. Overall, we conclude that the observed mtDNA phylogeny generally tracked the species phylogeny and was not consistent with widespread introgressions or horizontal gene transfer (HGT) of protein-coding genes across long evolutionary distances.

Analysis of mitochondrial genome content suggested that all mtDNAs likely retain the complete set of mitochondrially-encoded respiratory genes, including: the Complex IV components encoded by *COX1*, *COX2*, and *COX3*; the complex III component encoded by *COB*; and the ATP synthase components encoded by *ATP6*, *ATP8*, and *ATP9* (Figure 4A). The absence of some of these genes from a small number of assemblies was generally due to the assembly being fragmented or the annotation being manually removed due to issues with gene annotation (see Methods). In contrast, the mitochondrially-encoded components of the canonical Complex I (encoded by *NAD1-NAD6* and *NAD4L*) were surprisingly absent in several mtDNAs that otherwise appeared to be complete (Figure 4B). These genes are generally present in the mtDNAs of most fungi (Sandor et al., 2018) but were known to be absent in the orders Saccharomycetales and Saccharomycodales (Freel et al., 2015; Christinaki et al., 2022); indeed, our analysis is consistent with a major loss event in the common ancestor of these lineages. However, we also observed a single species lacking these genes in the order Dipodascales, *Nadsonia fulvescens* var. *fulvescens*, which is consistent with their absence in the related species *Nadsonia starkeyi-henricii* (O'Boyle et al., 2018) that was not included in this dataset, as well as a previously reported single-species loss event in the order Pichiales for *Ogataea philodendra* (Schikora-Tamarit et al., 2021). There were multiple independent losses within the order Phaffomycetales, including a single loss in the ancestor of *Candida ponderosae*, *Starmera amethionina*, and *Candida stellimalicola*, as well as potentially independent losses, including the previously reported loss for *Wickerhamomyces pijperi* (Schikora-Tamarit et al., 2021) and
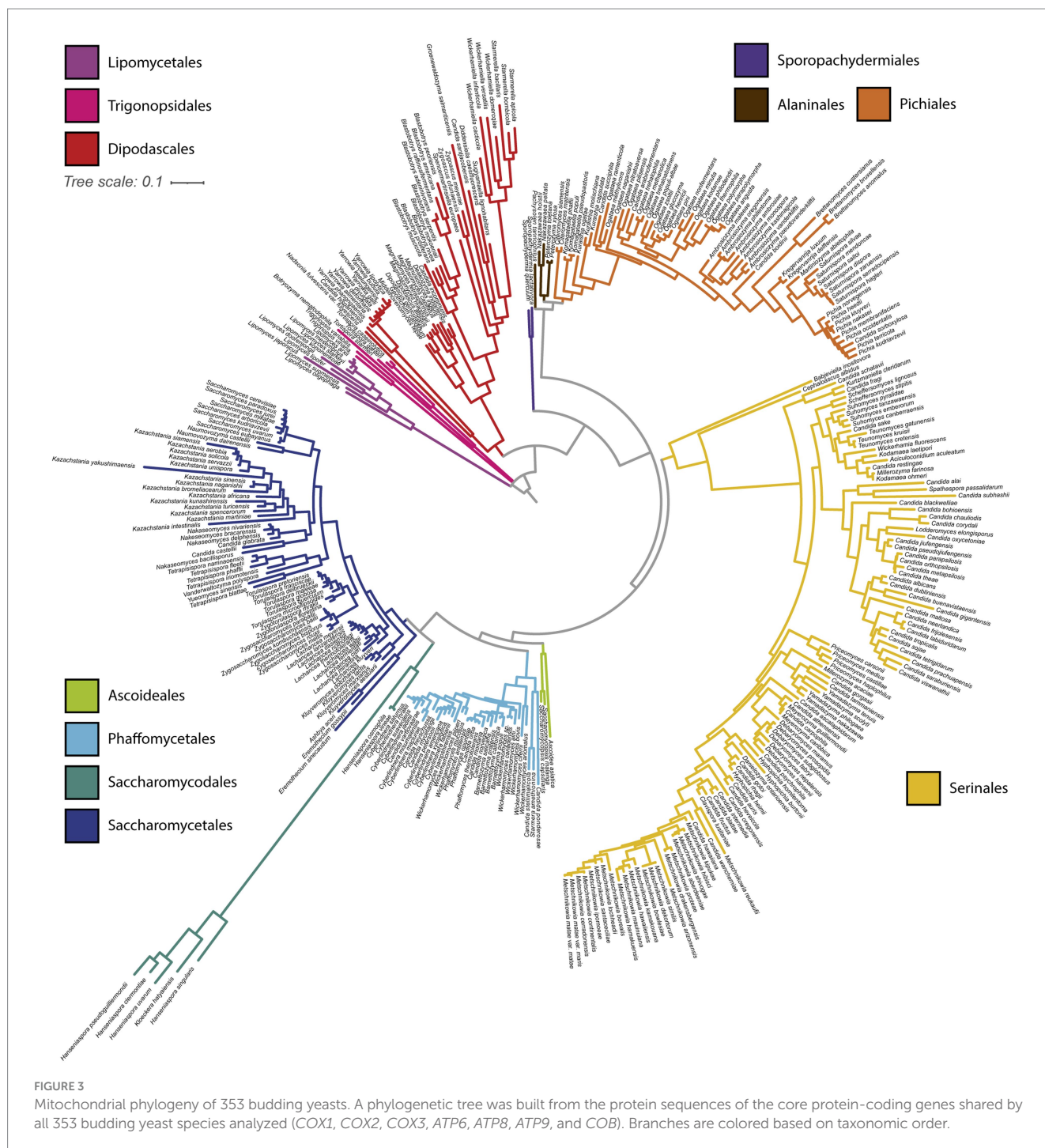
FIGURE 3
Mitochondrial phylogeny of 353 budding yeasts. A phylogenetic tree was built from the protein sequences of the core protein-coding genes shared by all 353 budding yeast species analyzed (*COX1, COX2, COX3, ATP6, ATP8, ATP9,* and *COB*). Branches are colored based on taxonomic order.

for *Cyberlindnera petersonii*. The distribution of the ribosomal protein encoded by *RPS3* was extremely patchy (Figure 4C). *RPS3* was not universally present in any taxonomic order, but all species in the dataset from the orders Serinales, Lipomycetales, and Sporopachydermiales lacked this gene. Difficulty in annotating this gene may have impeded its identification in some species (Bullerwell et al., 2000; Korovesi et al., 2018).

Despite similarities in gene content, genome size varied wildly at the extremes. *Pichia heedii* exceeded the previously largest observed Saccharomycotina mtDNA at 209,444 bp [versus the previous record of 187,024 bp in *Metschnikowia arizonensis* (Lee et al., 2020)], while

the smallest observed mtDNA was *Hanseniaspora pseudoguilliermondii* at 11,080 bp [versus the previous record of 18.8 kb in *Hanseniaspora uvarum* (Pramateftaki et al., 2006); Figure 4D]. The precise sizes of some mtDNAs were difficult to assess because not all assemblies were strictly complete, and short reads were not always capable of resolving genome structure reliably. The mtDNAs over 100 kb were typically more than double the size of any closely related species. Despite these observed extremes, the genome size of most species stayed within a range from approximately 20 to 80 kb (median size 39 kb, mean size 44 kb, standard deviation 23 kb). While this size variation is considerable in comparison with animal mtDNAs (Santamaria et al.,
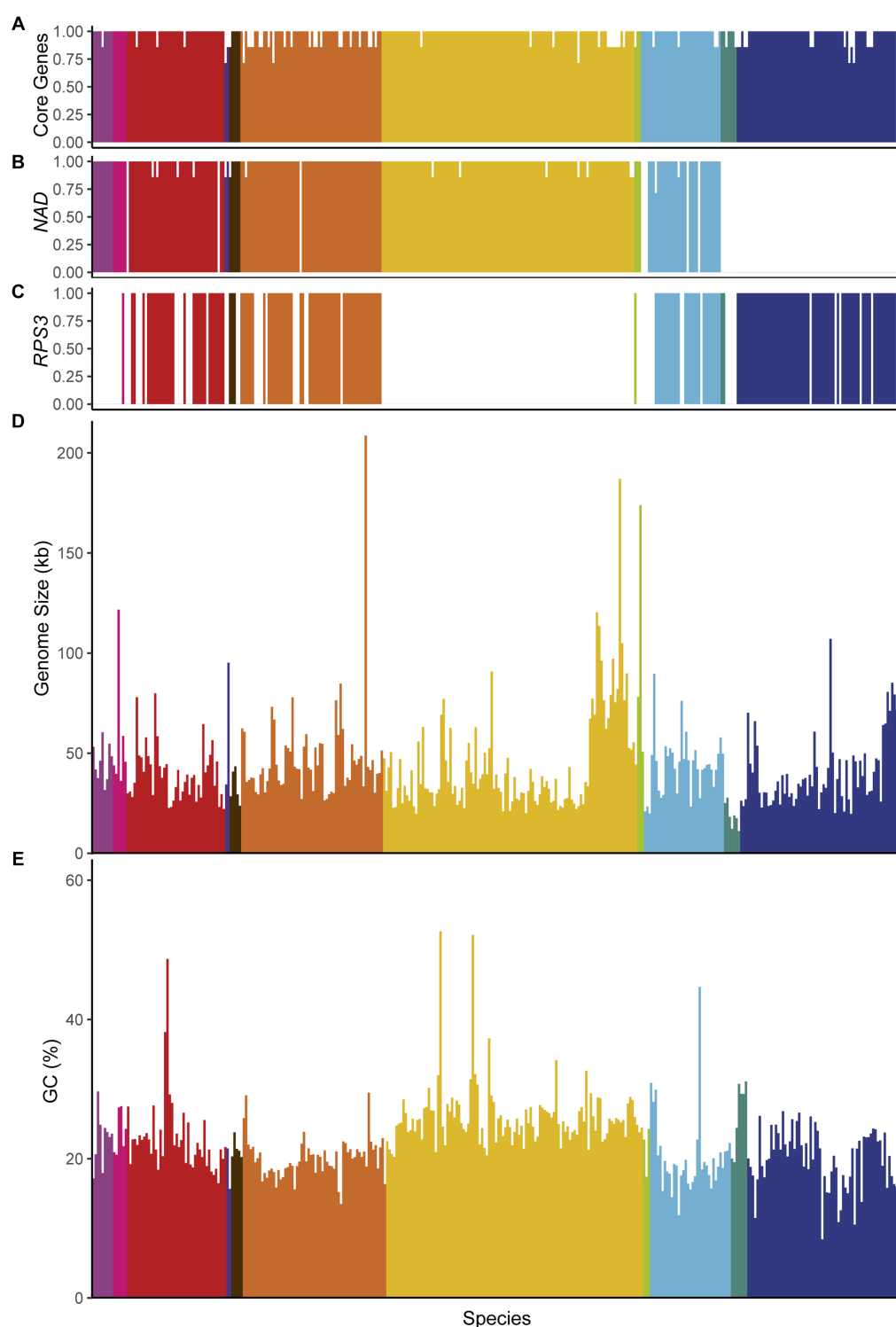
FIGURE 4
Genome characteristics. Genome characteristics are displayed and colored according to taxonomic order and placed based on position in the phylogenetic tree (left to right from Lipomycetales to Saccharomycetales, see Figure 3). The proportion of genes found in each genome are shown for: **(A)** core genes (*COX1, COX2, COX3, ATP6, ATP8, ATP9,* and *COB*), **(B)** Complex I genes (*NAD1-NAD6,* and *NAD4L*), and **(C)** the *RPS3* gene encoding a ribosomal protein. Genome sizes **(D)** and GC content **(E)** are indicated; both maintain a fairly limited range across the subphylum with a handful of extremes present across multiple taxonomic orders.

2007), it is within the ranges observed for other fungal mtDNAs (Sandor et al., 2018) and relatively low compared to plant mtDNAs (Gualberto and Newton, 2017).

The majority of species had similar GC content with a small number of outliers (Figure 4E). The average GC content was low (mean GC 22.5%, standard deviation 5.2%). Unusually high GC

contents were sporadically placed around the phylogeny, including *Candida subhashii* (52.7%) and *Candida gigantensis* (52.1%) in the order Serinales, *Magnusiomyces tetraspermus* (48.7%) in the order Dipodascales, and *Wickerhamomyces hampshirensis* (44.7%) in the order Phaffomycetales. The lowest value observed was for *Tetrapisispora blattae* at 8.4% (order Saccharomycetales), which was close to lowest value of 7.6% previously observed in *Saccharomycodes ludwigii* (Nguyen D. T. et al., 2020), which was not included in this dataset. Expansions of AT-rich intergenic regions have previously been reported to drive increases in genome size, which could drive a correlation between genome size and GC content. We found that, while this trend may be true in some groups, the overall correlation between genome size and GC content was poor and not significant after phylogenetic correction (r = −0.11, *p*-value 0.03; phylogenetically corrected r = −0.47, *p*-value 0.1, Supplementary Figure 2). Among the genomes over 100 kb, the average GC content (22.8%) was close to the global average. *Nakaseomyces bacillisporus* may have driven prior correlations within smaller scale analyses of the order Saccharomycetales (Xiao et al., 2017) due its unusually large size (107 kb) and low GC content (10. 9%), but this relationship does not appear to be strong across the expanded dataset. If expansions of intergenic regions drive size variation between distant species (Hao, 2022), then they likely do so in a manner that is GC-independent on average, although they may include biases in specific cases based on the base composition of the elements being expanded (e.g., AT-rich intergenic stretches in *Saccharomyces*).

2.3 Aerobic fermenters lack evidence for relaxed purifying selection

Metabolic strategies vary greatly among yeasts with regards to fermentation and respiration, which has been proposed to impact selection pressures on mitochondrial genes (Jiang et al., 2008). While many yeasts strongly respire fermentable carbon sources, such as glucose, there are many specialized yeasts, including most famously *S. cerevisiae*, that have developed metabolic strategies to preferentially ferment glucose and repress respiration, even in aerobic conditions (Merico et al., 2007; Rozpędowska et al., 2011; Hagman et al., 2013; Dashko et al., 2014; Hagman and Piškur, 2015). These aggressive fermenters are commonly said to exhibit Crabtree/Warburg Effect and are referred to as Crabtree/Warburg-positive (Diaz-Ruiz et al., 2011; Pfeiffer and Morley, 2014; Hammad et al., 2016). Given the relative disuse of respiration by this lifestyle, we hypothesized that the mitochondrially-encoded genes of Crabtree/Warburg-positive groups would exhibit elevated rates of non-synonymous substitutions due to relaxed purifying selection. Prior analysis of a limited set of species in the order Saccharomycetales had supported this model (Jiang et al., 2008).

To test the generality of this hypothesis, we determined the ratio of non-synonymous to synonymous substitution rates (ω) among groups at roughly the genus level (see Methods) across the phylogeny (Figure 5; Supplementary Table 2). We expected that ω would be highest in *Saccharomyces* and in related yeasts in the order Saccharomycetales that had undergone a whole-genome duplication (Marcet-Houben and Gabaldón, 2015; Wolfe, 2015) and were known to be strong fermenters, such as *Kazachstania* and *Nakaseomyces* (Hagman et al., 2013). Surprisingly, we observed that ω varied greatly

within taxonomic orders, with many groups exceeding the values observed for *Saccharomyces*. Indeed, the highest values were found in the order Dipodascales for yeasts in the *Wickerhamiella*/*Starmerella* clade and the grouping of yeasts most closely related to that clade (referred to as "Other Dipodascales" in Figure 5). The observed values for this clade are unlikely to be an artifact caused solely by long branch-lengths because the genus with the longest branch-lengths in the phylogeny (*Hanseniaspora* in the order Saccharomycodales) exhibited relatively moderate values. Within the order Saccharomycetales, we observed a general trend towards higher ω among yeasts that underwent the whole-genome duplication. The genus *Saccharomyces* followed this trend to some extent (genus mean ω 0.09 versus global mean 0.061), but this result was primarily driven by a single gene, *ATP8,* which had the highest value observed for all genes and groups and was driven by high values on the branches leading to *S. paradoxus and S. arboricola* (0.355). When this gene was excluded, the remaining genes defied the trend (0.046). *ATP8* is highly conserved between *S. cerevisiae* strains (Wolters et al., 2015), which suggests inter- and intra-specific patterns of variation can differ greatly. Given the high ω values for many yeasts not known to be Crabtree/Warburg-positive and the relatively low ω for most *Saccharomyces* genes, we conclude that our much-expanded dataset does not support the previously proposed model of pervasive relaxed purifying selection on the mitochondrially-encoded genes of aerobic fermenters.

2.4 Evidence for horizontal transfer of mitochondrial introns across orders

Mitochondrial introns vary widely in yeasts, largely due to sporadic gains and losses (Xiao et al., 2017). Intron-encoded homing endonucleases are thought to drive intron turnover and potentially HGT of introns between species (Lang et al., 2007; Wu and Hao, 2014). The highest numbers of introns were observed in *Magnusiomyces* (mean 18 introns per species versus global mean 5.4; Supplementary Table 3), *Metschnikowia* (10.7), and *Yarrowia* (10.5, including other closely related anamorphic species that have yet to be reassigned to this genus). The lowest values were observed in *Eremothecium* (0.33) and *Deakozyma* (0.5), both of which included species that were completely free of introns. Nearly all introns were encoded within *COX1* (55.3%), *COB* (30.1%), or *NAD5* (7.4%); the remaining genes had <2% each. The small range of gene targets is consistent with intron homing by endonucleases transferring introns, including by HGT, to a limited range of target sites.

We identified potential intron HGTs based on BLAST comparisons of all mitochondrial introns observed using a conservative threshold to classify introns as unique, shared within a group (identical groupings as for the selection analysis above), shared within and between groups, or solely between groups (>50% of maximum possible bit score, Figure 6A). Most introns observed did not share high sequence similarity to introns from other species (65.6%), while most of the remainder were shared within a group (30%). A small number were shared across groups, and this phenomenon was especially common in the order Saccharomycetales (Figure 6B). Clustering the introns based on pairwise BLAST hits generated 271 clusters of related introns (Supplementary Table 3). In all clusters spanning multiple groups or orders, the shared introns
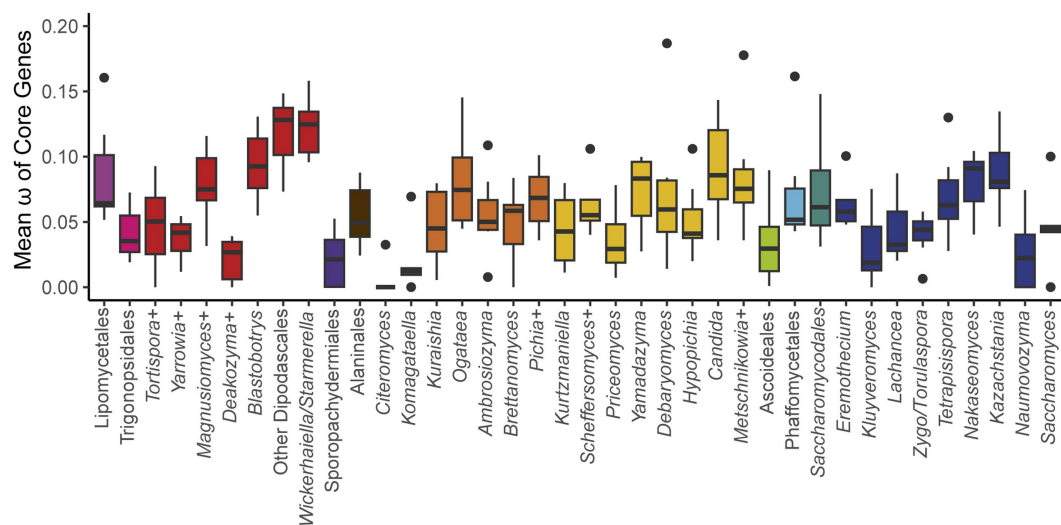
FIGURE 5
Mean ω of core genes. The ratio of non-synonymous to synonymous substitution rates for each of the core protein–coding genes was calculated for groups across the phylogeny (+ indicates that additional closely related species that are not currently classified in that genus were included, see Supplementary Table 1). The box and whisker plots show the distribution of ω among genes within each group (boxes centered at median encompassing the interquartile range, whiskers up to 1.5 times the interquartile range, and outlier genes shown as individual datapoints). Two extreme outlier genes were omitted from the graph: *ATP8* for *Saccharomyces* (0.355) and *COB* for *Kurtzmaniella* (0.250). Groups with aerobic fermenters, such as *Saccharomyces*, *Kazachstania*, and *Nakaseomyces*, do not exhibit significantly elevated ratios relative to the rest of the subphylum.

were of the same variety (group I or II), and nearly all introns encoded an open reading frame (ORF) encoding a putative homing endonuclease or reverse-transcriptase gene of the same motif (either LAGLIDADG or GIY-YIG). Rarely, an intron within the community had an ORF whose motif was not recognizable or, even more rarely, lacked an ORF altogether. Our results are consistent with prior findings of intron mobility being primarily driven by intronic ORFs,which may then be lost long before the intron, resulting in poor conservation within groups and a high prevalence of HGT (Wu et al., 2015).

Only a single cluster contained introns that were found within different genes due to homology between *Metschnikowia mauinuiana* *NAD2* intron 1 and *COX1* intron 1 from the same species and from *Metschnikowia hawaiiensis* (Figure 6C). *NAD2* is duplicated in *M. mauinuiana*, but only one copy has been colonized by this intron; however, the second copy contains a 560-bp duplication identical to the 3′ end the intron. Thus, *M. mauinuiana NAD2* intron 1 may be misannotated and may instead be a 3′ terminal element that could be translated as an extension of the upstream gene; a similar phenomenon has been observed for *COX2* and other genes in *Saccharomyces* (Peris et al., 2017). *M. mauinuiana COX1* intron 1 had homology to the reverse transcriptase encoded in intron 1 of *S. cerevisiae COX1*; however, *M. mauinuiana NAD2* intron 1 appeared to be truncated, which disrupts the intronic open reading frame. Thus, *M. mauinuiana NAD2* intron 1 may better be thought of as an example of how a 3′ terminal element may be formed by an intronic mobile element acquiring a novel insertion site. The high number of introns in these species may be increasing the odds of such events in this genus, which has been speculated to have the strangest mitochondrial genomes (Lee et al., 2020).

We observed 22 clusters that contained introns spanning multiple groups, including four that contained introns spanning multiple

orders (Figure 6D; Supplementary Figure 3); these clusters are the top candidates for HGT events in our dataset. For example, the fifth intron of *COX1* from *S. cerevisiae* (sometimes referred to as aI5α) shared homology with several *Saccharomyces COX1* introns, as well as *Hanseniaspora vineae COX1* intron 3 (Figure 6D). This cluster of introns may also include *Lachancea kluyveri COX1* intron 5, but this connection was only supported for *Saccharomyces jurei COX1* intron 4 (Figure 6D). The introns in the clusters were all Group I with an ORF of the LAGLIDADG motif with two exceptions: (1) the *H. vineae* intron, in which the ORF was annotated but the motif was not; and (2) the *S. mikatae* intron, which lacked a clear ORF and was of indeterminate type because of possible degeneration (Supplementary Table 3). All other *H. vineae COX1* introns (order Saccharomycodales) shared limited homology to introns within the order Saccharomycetales, but it was well below our cutoff; since they shared no clear homology to other *Hanseniaspora* introns, these are also candidates for HGT, albeit more tentative ones. Two of the four clusters with evidence of cross-order HGT involved introns from *Hypophichia burtonii*, which suggests that this species may contain several highly active intronic mobile elements. Interestingly, this lineage also appears to have been an HGT donor of nuclear-encoded genes for utilization of the sugar galactose (Haase et al., 2021). We conclude that homology in homing endonuclease target sites likely enables the HGT of these selfish elements, even across large phylogenetic distances, at least in rare cases.

3 Discussion

As high-throughput sequencing revolutionized genomics, advances in yeast mitochondrial genomics were initially delayed. Early high-throughput datasets generated only partial sequences, potentially due
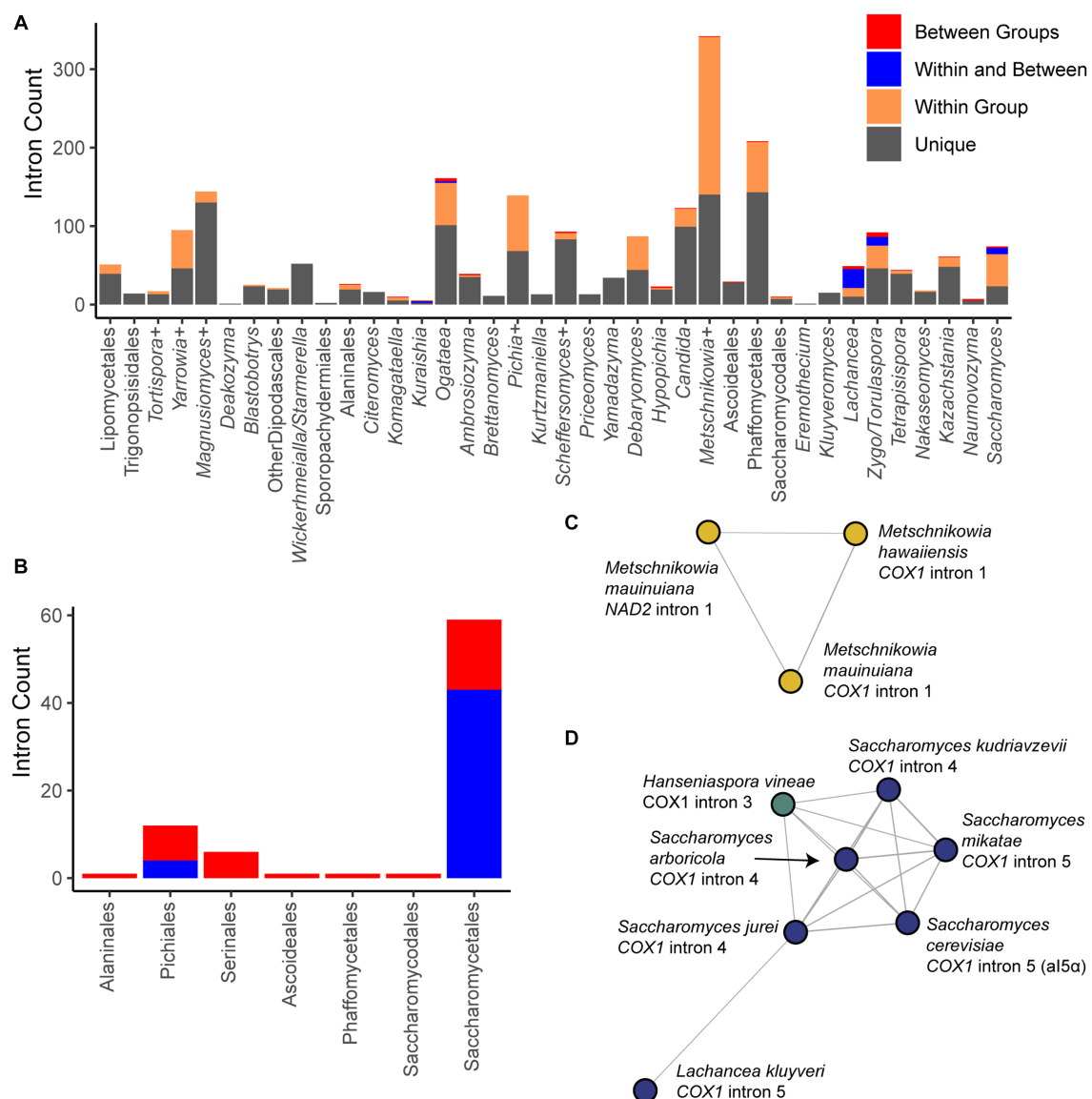
FIGURE 6
Intron diversity. **(A)** Introns were classified based on pairwise BLAST hits as unique to that species, present in multiple species of the group, shared within and between groups, or only between groups. The counts of introns in each category within each group are displayed. **(B)** The counts of introns in each taxonomic order that were shared or found only between groups are displayed. Orders not listed had no introns in these categories. **(C)** Introns were clustered based on shared BLAST hits, and the single cluster containing hits shared across multiple genes is displayed. Nodes are colored based on taxonomic order as in Figure 2 (all Serinales). **(D)** A cluster of introns is displayed that spans the orders Saccharomycetales and Saccharomycodales, including *Saccharomyces* spp., *Lachancea kluyveri*, and *Hanseniaspora vineae*. Nodes are colored based on taxonomic order as in Figure 3.

to biases against AT-rich sequences (Chen et al., 2013; Ross et al., 2013). Advances in methodology led to large numbers of *S. cerevisiae* mitochondrial genomes being sequenced in tandem with their nuclear genomes (Strope et al., 2015). More recently, even very large population datasets produced mitochondrial genomes concurrently with the nuclear genomes (De Chiara et al., 2020). Prior to this study, these advances had not yet come to bear for large species-rich datasets, with targeted post-hoc searches of published assemblies yielding limited numbers of additional mtDNAs (Christinaki et al., 2022). Here, we have demonstrated that, even for short-read-only datasets, it is possible to extract high-quality mitochondrial genomes with a high success rate from datasets originally collected for nuclear sequencing. As yeast

genomics progresses further, the mitochondrial component need not be an afterthought.

Despite these advances, limitations remain. Mitochondrial genome structure is complex and not always readily solvable through short reads alone. For example, the *S. cerevisiae* mtDNA maps genetically as circular, but the predominant molecular form is a linear concatemer of multiple genome units (Maleszka et al., 1991; Solieri, 2010). Other species exhibit monomeric linear forms (Valach et al., 2011), including *C. albicans* and related species (Nosek et al., 1995; Gerhold et al., 2010), or even have capping terminal inverted repeats as seen in *H. uvarum* (Pramateftaki et al., 2006) and several Phaffomycetales species (Dinouël et al., 1993). The linearization in

Hanseniaspora may be relatively recent, as another species within the order Saccharomycodales but outside the genus, *Saccharomycodes ludwigii,* maps as a circle (Nguyen D. T. et al., 2020). The exact transition point is difficult to assess as many of the *Hanseniaspora* assemblies did not cleanly assemble the repeat elements, although there are some indications it may have occurred during the formation of the faster-evolving lineage (Steenwyk et al., 2019). Similar linear genomes with capping structures have been observed in other fungi (Forget et al., 2002; van de Vossenberg et al., 2018), but the evolutionary pressures leading to linearization are not well understood, although mobile elements have been hypothesized as the origin of capping elements (Nosek and Tomáška, 2003). Long-read sequencing technologies are a promising avenue to obtain not only complete mtDNAs, which short reads alone failed to provide for many species, but also to resolve complex genome structures by generating reads longer than a single genome unit in length. This strategy has already been successful at investigating large-scale deletion mutations in *S. cerevisiae* (Nunn and Goyal, 2022). However, specialized assemblers, similar in principle to those used here for reassembly of the short reads, will be needed because current long-read assemblers, such as canu (Koren et al., 2017), frequently misassemble circular-mapping genomes (Wick and Holt, 2019).

The most striking variation seen among the mtDNAs is the complete loss of canonical Complex I in Saccharomycetales, Saccharomycodales, and several additional lineages across the phylogeny. In *S. cerevisiae*, the acquisition of genes encoding a multi-unit alternative NADH:ubiquinone oxidoreductase facilitated this loss (Luttik et al., 1998; Kerscher, 2000), albeit at the cost of a loss in potential proton motive force. The mechanisms that allowed for this loss in the other independent events are currently unclear, but they suggest that multiple species may also have potentiating factors that could facilitate loss. Canonical Complex I is encoded by both nuclear and mitochondrial genes, but these nuclear genes were concomitantly lost in *S. cerevisiae* with the mitochondrial genes. If the same pattern persists across independent loss events, then any uncharacterized genes that were also lost in tandem may also play a role in Complex I function. The best available hypothesis explaining Complex I loss suggests that it is an adaptation to mitigate reactive oxygen species, and indeed, it has been observed that nuclear genes experiencing duplication, loss, or altered selection were specifically enriched for genes involved in oxidation–reduction processes (Schikora-Tamarit et al., 2021).

Originally, we hypothesized that preference for aerobic fermentation would be a major factor driving mitochondrial genome variation. Previously, it had even been hypothesized to play a significant role in the loss of Complex I as *Brettanomyces* species were the only others known to be Crabtree/Warburg-positive but still encode a canonical Complex I (Freel et al., 2015). Given that multiple losses of Complex I were observed in species not known to be Crabtree/Warburg-positive and given the lack of evidence for relaxed purifying selection in aerobic fermenters, it is not evident that this shift in metabolism is a major driver of mitochondrial gene evolution. An important caveat is that the methodology employed here may be limited by current datasets on the distribution of aerobic fermentation, which extrapolate from only a handful of well-characterized species. For example, the *Wickerhamiella*/*Starmerella* clade merits further attention due to the high rates of non-synonymous variation observed and potential environmental preferences for

sugar-rich environments in this group (Gonçalves et al., 2020). Additionally, estimating selection at the group level may obscure patterns of selection that vary more between closely related species than between groups, as previously observed for *Lachancea* species (Freel et al., 2014). Focusing on selection pressures at the level of individual genes may also be more illuminating. The ω rates varied more for comparisons for the same gene across groups (mean variance 0.0018) than for comparisons of different genes within groups (0.0014), which is contrary to the expectation that a given gene should face similar functional, constraints regardless of group, and thus have a smaller variance. For example, while *ATP9* is the most conserved gene within *Saccharomyces*, it is the least conserved in *Nakaseomyces*. If aerobic fermentation does play a role, it may relax selective pressure on some genes but increase purifying selection for others.

Several alternative hypotheses merit further exploration to explain the observed variation in evolutionary rates. Demographic effects are well known to generate patterns reminiscent of selective pressures (Johri et al., 2020), but the ecologies and population histories of these species are not well known enough to determine if they have played a role here. Mutational pressures from reactive oxygen species could also cause biases in certain species for which the production of or exposure to free radicals is elevated (Melvin and Ballard, 2017). The best explanation for variation in the mitochondrially-encoded proteins may actually be in the nuclear genome. The mitochondrially-encoded genes rely on nuclear-encoded proteins for their replication, transcription, translation, assembly, and formation of the respiratory complexes with mostly nuclear-encoded subunits (Lipinski et al., 2010; Burton and Barreto, 2012). The high degree of mitochondrial-nuclear epistasis has been observed to generate higher evolutionary rates in interacting nuclear proteins in a well-established copepod model (Barreto et al., 2018). Further analyses are needed to determine whether other selection or mutational pressures at the nuclear level may cause accelerated mitochondrial gene evolution independent of aerobic fermentation.

Mitochondrial introns may also serve an important role in shaping mitochondrial gene evolution. Homing endonucleases, which are encoded within mitochondrial introns or in downstream open reading frames at the 3′ end of mitochondrial genes, have been shown to modify sequences adjacent to the insertion site (Repar and Warnecke, 2017; Xiao et al., 2017; Wu and Hao, 2019). Transfers between groups, and potentially between orders, may introduce non-synonymous variation due to co-conversion of flanking sequences during insertion. We observed a large proportion of unique introns in our dataset, which is consistent with high rates of intron turnover underlying presence/absence variation. However, we have likely underestimated the true proportion of introns shared within groups due to the stringent criteria applied and the rapid decay of detectable sequence homology due to high mtDNA mutation rates (Sharp et al., 2018). Mitochondrial introns have been known to jump between different kingdoms between the symbiotic components of lichens (Mukhopadhyay and Hausner, 2021). Certain ecological conditions, such as coculture of *Saccharomyces* and *Hanseniaspora* during wine fermentation (Langenberg et al., 2017), may similarly facilitate horizontal transfer.

The mitochondrial genomes generated in this study provide many opportunities to further our understanding of evolution beyond the scope of this study. Pairing the data with the previously generated nuclear genomes will help elucidate the interplay between these two

genomes. Mitochondrial-nuclear epistasis has been demonstrated to affect phenotypic variation in yeasts (Paliwal et al., 2014; Nguyen T. H. M. et al., 2020; Visinoni and Delneri, 2022; Biot-Pelletier et al., 2023; Nguyen et al., 2023) and a diverse array of model systems (Dowling et al., 2007; Burton and Barreto, 2012; Mossman et al., 2016). For many existing mitochondrial genomes, any analysis of such interactions was previously often complicated by the lack of a corresponding nuclear genome or by mismatches between the strains sequenced for a given species. By mining most of this new mtDNA dataset from a dataset of high-quality nuclear genomes (Shen et al., 2018), many of these previous limitations have been lifted, which has already enabled the novel insights described here. The breadth and the richness of these paired nuclear-mitochondrial datasets promise to greatly accelerate research into the evolution of yeast mitochondrial genomes.

4 Materials and methods

4.1 Mitochondrial genome mining, assembly, and annotation

We searched 332 yeast genome assemblies for mitochondrial contigs using a two-pronged, reference-based approach (Shen et al., 2018). First, we curated a set of reference mtDNAs from all accessions in Genbank matching Saccharomycotina and with the source as "mitochondrion" in September 2018 to generate a set of 110 published mtDNAs with a single representative per species (Supplementary Table 1). Existing annotations were curated based on length and presence of stop codons, and they were renamed for consistent formatting. When annotations were not available, new annotations were generated using MFANNOT (Lang et al., 2007). We identified putative mitochondrial contigs based on two BLAST strategies searches (v2.8.1). First, the coding sequences (CDS) from the curated references were used as queries to search each assembly, and contigs with at least 10 hits >70% coverage and e-value <0.001 were retained. Second, the contigs from each assembly were used as queries against the complete reference mtDNAs, and contigs with at least one 25% coverage hit with e-value <0.001 were retained. These contigs were then preliminarily annotated using MFANNOT (Lang et al., 2007) to estimate gene content (Lang et al., 2007). To eliminate contigs that were likely short duplicates of mitochondrial sequences transferred to the nuclear genome, also known as NUMTs (Hazkani-Covo et al., 2010; Xue et al., 2023), we filtered out contigs that did not possess at least one mitochondrial gene per 20 kb. Contigs larger than 300 kb were also removed to eliminate any complete mtDNA duplicates in large nuclear contigs.

Assembly methods for nuclear genomes are generally not optimized for mitochondrial sequences, so we reassembled genomes for which sequencing reads were readily available, including 196 species sequenced in Shen et al., 2018 and 92 additional species included in that dataset that we resequenced to replace an older nuclear assembly as part of the Y1000+ Project (Supplementary Table 1; Opulente et al., 2023). Reassembly was done using either plasmidSPAdes v3.9.0 (Antipov et al., 2016) or NOVOPlasty v4.2 (Dierckxsens et al., 2017). We annotated these assemblies using MFANNOT (Lang et al., 2007) and then searched for mitochondrial contigs as described above. For NOVOPlasty, multiple assemblies

were constructed using different seeds either using the genes found in the putative mitochondrial contigs extracted from the nuclear assembly or the CDS from the closest available genome based on the nuclear phylogeny in the curated reference set. The putative mitochondrial contigs isolated from the nuclear assembly and the mitochondrial reassemblies were assessed based on completeness (% of expected genes present, excluding *RPS3* and Complex I genes when none were present), contiguity (% of genes found on each contig), and circularity (count of reads that map across the contig endpoints after shifting the sequence such that the original breakpoint is internal in the permuted contig), and a single assembly was chosen for each species. We prioritized completeness and used contiguity and circularity to break ties. Generally, NOVOPlasty performed best, followed by plasmidSPAdes, while the existing contigs from the nuclear assembly were best in a small minority of cases. For the final dataset, we combined these assemblies with the curated reference set, retaining one assembly per species and choosing the reference assembly for a species when available.

All new assemblies, as well as existing mtDNAs that were not annotated, were annotated from scratch; all genome annotations, including published ones, were curated for consistency and to improve accuracy as described below. The translation table for each species was estimated using codetta v2.0 (Shulgina and Eddy, 2021). Yeast mitochondrial translation tables fall into either the Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma/Spiroplasma Code (NCBI table 4, hereafter referred to as the fungal code), which is consistent with other fungi, or the yeast mitochondrial code (NCBI table 3, originally based on *S. cerevisiae,* hereafter referred to as the *Saccharomyces* code) based on additional reassignments of AUA and CUN codons, which typically define the order Saccharomycetales. The exact placement of this transition was difficult to determine due to a loss of CUN codons in many Saccharomycetales, particularly *Kluyveromyces* species and other closely related genera. In many species, the CUN reassignment is supported by codetta, but the AUA reassignment is not, and the modified tRNA required for this reassignment is not present, which is consistent with a previous analysis of codon usage among Saccharomycotina mtDNAs (Christinaki et al., 2022). Currently, no translation table exists for the CUN reassignment without the AUA reassignments, so we used the *Saccharomyces* code when the CUN reassignment was supported and the fungal code for all others (Supplementary Table 1). The AUA reassignment in the *Saccharomyces* code allows for this codon to be treated as a start codon by MFANNOT, which resulted in many misannotations at the 5′ end of genes. No examples of AUA being used as a valid start codon in yeasts have been described. We rectified this issue by reannotating all assemblies using table 4 to define start and end coordinates; we then used the *Saccharomyces* code for translation when appropriate. Finally, all annotations (for new and existing mtDNAs) were further manually curated to eliminate truncated genes, annotations split across contigs, and annotations containing large extensions due to misannotated introns or readthroughs. Our analysis does not fully account for the full range of alternative translation tables previously reported in yeasts, and we did not detect some previously reported reassignments in *Eremothecium* species (Ling et al., 2014; Noutahi et al., 2017).We identified several *Kazachstania* species with frameshifts consistent with the +1C frameshift mechanism previously described (Szabóová et al., 2018). To match the formatting in GenBank for those references, we encoded these as single-bp introns, but

these were excluded from all intron analyses. We did not observe any *byp* elements, as described in *Magnusiomyces capitatus,* in the coding sequences of other species using methods previously described (Lang et al., 2014). Assemblies and annotations are available through GenBank (Bioproject PRJNA998421) and in additional formats in the figshare repository (see Data Availability Statement).

4.2 Mitochondrial phylogeny construction

We determined phylogenetic relationships among mitochondrial genomes based on the core set of genes shared by all species: *COX1*, *COX2*, *COX3*, *COB*, *ATP6*, *ATP8*, and *ATP9*. Complex I genes were excluded due their loss in a large fraction of the species. Protein sequences were aligned for each gene using MAFFT using the E-INS-I option (Katoh and Standley, 2013), and CDS were codon-aligned using the protein alignment. The alignments were concatenated and then filtered to retain only sites in which 95% of sequences were not gaps using trimAl (Capella-Gutiérrez et al., 2009). We built multiple phylogenies from the filtered alignment using IQ-TREE using the mitochondrial substitution model (Minh et al., 2020). These phylogenies were highly concordant, except for the placement of the fast-evolving *Hanseniaspora* lineage. The topology most consistent with the nuclear phylogeny was selected as the final tree. Phylogenetic correction of correlations of genome size versus GC content were done using a generalized least squares approach [using gls from nlme (Pinheiro and Bates, 2023)] using a co-variation matrix generated using a Brownian motion model [using corPagel from ape (Paradis and Schliep, 2019)]. Annotation files for the phylogenetic tree to display various features using the iTOL tree viewer (Letunic and Bork, 2021) are included in the figshare repository (see Data Availability Statement). Putative mitochondrial contigs with no annotations for respiratory subunits were removed from GenBank submission but are included in the figshare.

4.3 Estimating patterns of selection

To investigate patterns of selection on mitochondrial genes, we split the phylogeny into smaller groups at roughly the genus level to avoid saturation of synonymous substitutions (Supplementary Table 1). For each of the genes in the core set, we built subtrees for each group and estimated ω along each branch of the subtree using PAML under model 1 (allowing variable ω for each branch; Yang, 2007). For each gene, the ω value was determined as the mean of the values for all branches in the subtree for which there were sufficient synonymous substitutions (dS > 0.01).

4.4 Evaluating evidence for horizontal transfer of mitochondrial introns

Possible HGTs of mitochondrial introns were determined based on an all-versus-all BLAST of mitochondrial introns against each other. Mitochondrial introns among closely related species are expected to share limited sequence similarity due to poor conservation of non-coding sequences, though elements that contribute to intron splicing may be under purifying selection. Thus, we set a conservative

threshold that the bit score of each hit must be at least 50% of the maximum possible bit score determined by the self-to-self comparison of each intron and have an e-value $<10^{-10}$. Shared relationships within groups are likely to be due to vertical descent, although there is evidence that HGT frequently occurs at this scale (Wu and Hao, 2014), but such high sequence similarity at large phylogenetic distances is likely due to HGT. Clustering of intron sequences was performed using the Louvain method (Blondel et al., 2008) implemented in the igraph (Csárdi et al., 2023) package of R.

Data availability statement

All supporting data and analyses are available at figshare doi: https://doi.org/10.6084/m9.figshare.23799789. The genomes have been deposited in Genbank (Bioproject PRJNA998421).

Author contributions

Funding

Acknowledgments

services; Wisconsin Energy Institute staff for computational support; and the Center for High-Throughput Computing at the University of Wisconsin-Madison (https://doi.org/10.21231/GNT1-HW21).

Conflict of interest

AR is a scientific consultant for LifeMine Therapeutics, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1268944/full#supplementary-material

References

Adams, K. L., and Palmer, J. D. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* 29, 380–395. doi: 10.1016/S1055-7903(03)00194-5

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. A. (2016). PlasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32, 3380–3387. doi: 10.1093/bioinformatics/btw493

Barreto, F. S., Watson, E. T., Lima, T. G., Willett, C. S., Edmands, S., Li, W., et al. (2018). Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat. Ecol. Evol.* 2, 1250–1257. doi: 10.1038/s41559-018-0588-1

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* 21, 163–193. doi: 10.1111/j.1096-0031.2005.00059.x

Biot-Pelletier, D., Bettinazzi, S., Gagnon-Arsenault, I., Dubé, A. K., Bédard, C., Nguyen, T. H. M., et al. (2023). Evolutionary trajectories are contingent on mitonuclear interactions. *Mol. Biol. Evol.* 40, 1–16. doi: 10.1093/molbev/msad061

Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008:P10008. doi: 10.1088/1742-5468/2008/10/P10008

Bullerwell, C. E., Burger, G., and Lang, B. F. (2000). A novel motif for identifying Rps3 homologs in fungal mitochondrial genomes. *Trends Biochem. Sci.* 25, 363–365. doi: 10.1016/S0968-0004(00)01612-1

Burton, R. S., and Barreto, F. S. (2012). A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol. Ecol.* 21, 4942–4957. doi: 10.1111/mec.12006

Butler, G., Rasmussen, M. D., Lin, M. F., Santos, M. A. S., Sakthikumar, S., Munro, C. A., et al. (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459, 657–662. doi: 10.1038/nature08064

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., and Hwang, C.-C. (2013). Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* 8:e62856. doi: 10.1371/journal.pone.0062856

Christinaki, A. C., Kanellopoulos, S. G., Kortsinoglou, A. M., Andrikopoulos, M., Theelen, B., Boekhout, T., et al. (2022). Mitogenomics and mitochondrial gene phylogeny decipher the evolution of Saccharomycotina yeasts. *Genome Biol. Evol.* 14, 1–19. doi: 10.1093/gbe/evac073

Csárdi, G., Nepusz, T., Müller, K., Horvát, S., Traag, V., Zanini, F., et al. (2023). igraph for R: R interface of the igraph library for graph theory and network analysis. Available at: https://zenodo.org/record/8046777 (Accessed July, 2023).

Dashko, S., Zhou, N., Compagno, C., and Piskur, J. (2014). Why, when, and how did yeast evolve alcoholic fermentation? *FEMS Yeast Res.* 14, 826–832. doi: 10.1111/1567-1364.12161

De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., and Liti, G. (2020). Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biol.* 18, 49–15. doi: 10.1186/s12915-020-00786-4

Diaz-Ruiz, R., Rigoulet, M., and Devin, A. (2011). The Warburg and Crabtree effects: on the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochim. Biophys. Acta Bioenerg.* 1807, 568–576. doi: 10.1016/j.bbabio.2010.08.010

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18. doi: 10.1093/nar/gkw955

Dinouël, N., Drissi, R., Miyakawa, I., Sor, F., Rousset, S., and Fukuhara, H. (1993). Linear mitochondrial DNAs of yeasts: closed-loop structure of the termini and possible linear-circular conversion mechanisms. *Mol. Cell. Biol.* 13, 2315–2323. doi: 10.1128/mcb.13.4.2315

Dowling, D. K., Abiega, K. C., and Arnqvist, G. (2007). Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. *Evolution* 61, 194–201. doi: 10.1111/j.1558-5646.2007.00016.x

Dowling, D. K., and Wolff, J. N. (2023). Evolutionary genetics of the mitochondrial genome: insights from *Drosophila*. *Genetics* 224: iyad036. doi: 10.1093/genetics/iyad036

Forget, L., Ustinova, J., Wang, Z., Huss, V. A. R., and Lang, B. F. (2002). *Hyaloraphidium curvatum*: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Mol. Biol. Evol.* 19, 310–319. doi: 10.1093/oxfordjournals.molbev.a004084

Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B. (1998). The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* 440, 325–331. doi: 10.1016/S0014-5793(98)01467-7

Freel, K. C., Friedrich, A., Hou, J., and Schacherer, J. (2014). Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. *Genome Biol. Evol.* 6, 2586–2594. doi: 10.1093/gbe/evu203

Freel, K. C., Friedrich, A., and Schacherer, J. (2015). Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.* 15, 1–9. doi: 10.1093/femsyr/fov023

Gerhold, J. M., Aun, A., Sedman, T., Jõers, P., and Sedman, J. (2010). Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol. Cell* 39, 851–861. doi: 10.1016/j.molcel.2010.09.002

Gonçalves, P., Gonçalves, C., Brito, P. H., and Sampaio, J. P. (2020). The *Wickerhamiella/Starmerella* clade—a treasure trove for the study of the evolution of yeast metabolism. *Yeast* 37, 313–320. doi: 10.1002/yea.3463

Groenewald, M., Hittinger, C. T., Bensch, K., Opulente, D. A., Shen, X.-X., Li, Y., et al. (2023). A genome-informed higher rank classification of the biotechnologically important fungal subphylum Saccharomycotina. *Stud. Mycol.* 105, 1–22. doi: 10.3114/sim.2023.105.01

Gualberto, J. M., Mileshina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016

Gualberto, J. M., and Newton, K. J. (2017). Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* 68, 225–252. doi: 10.1146/annurev-arplant-043015-112232

Haase, M. A. B., Kominek, J., Opulente, D. A., Shen, X. X., LaBella, A. L., Zhou, X., et al. (2021). Repeated horizontal gene transfer of *GAL*actose metabolism genes violates Dollo's law of irreversible loss. *Genetics* 217:iyaa012. doi: 10.1093/GENETICS/IYAA012

Hagman, A., and Piškur, J. (2015). A study on the fundamental mechanism and the evolutionary driving forces behind aerobic fermentation in yeast. *PLoS One* 10, 1–24. doi: 10.1371/journal.pone.0116942

Hagman, A., Sall, T., Compagno, C., and Piskur, J. (2013). Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS One* 8:e68734. doi: 10.1371/journal.pone.0068734

Hammad, N., Rosas-Lemus, M., Uribe-Carvajal, S., Rigoulet, M., and Devin, A. (2016). The Crabtree and Warburg effects: do metabolite-induced regulations participate in their induction? *Biochim. Biophys. Acta Bioenerg.* 1857, 1139–1146. doi: 10.1016/j.bbabio.2016.03.034

Hao, W. (2022). From genome variation to molecular mechanisms: what we have learned from yeast mitochondrial genomes? *Front. Microbiol.* 13, 1–8. doi: 10.3389/fmicb.2022.806575

Hazkani-Covo, E., Zeller, R. M., and Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 6:e1000834. doi: 10.1371/journal.pgen.1000834

Hittinger, C. T. (2013). Saccharomyces diversity and evolution: a budding model genus. *Trends Genet.* 29, 309–317. doi: 10.1016/j.tig.2013.01.002

Hittinger, C. T., Rokas, A., Bai, F. Y., Boekhout, T., Gonçalves, P., Jeffries, T. W., et al. (2015). Genomics and the making of yeast biodiversity. *Curr. Opin. Genet. Dev.* 35, 100–109. doi: 10.1016/j.gde.2015.10.008

Jiang, H., Guan, W., Pinney, D., Wang, W., and Gu, Z. (2008). Relaxation of yeast mitochondrial functions after whole-genome duplication. *Genome Res.* 18, 1466–1471. doi: 10.1101/gr.074674.107

Johnston, I. G., and Williams, B. P. (2016). Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. *Cell Syst.* 2, 101–111. doi: 10.1016/j.cels.2016.01.013

Johri, P., Charlesworth, B., and Jensen, J. D. (2020). Toward an evolutionarily appropriate null model: jointly inferring demography and purifying selection. *Genetics* 215, 173–192. doi: 10.1534/genetics.119.303002

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kerscher, S. J. (2000). Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochim. Biophys. Acta Bioenerg.* 1459, 274–283. doi: 10.1016/S0005-2728(00)00162-6

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Korovesi, A. G., Ntertilis, M., and Kouvelis, V. N. (2018). Mt-rps3 is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. *Mol. Phylogenet. Evol.* 127, 74–86. doi: 10.1016/j.ympev.2018.04.037

Lang, B. F., Jakubkova, M., Hegedusova, E., Daoud, R., Forget, L., Brejova, B., et al. (2014). Massive programmed translational jumping in mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5926–5931. doi: 10.1073/pnas.1322190111

Lang, B. F., Laforest, M.-J., and Burger, G. (2007). Mitochondrial introns: a critical view. *Trends Genet.* 23, 119–125. doi: 10.1016/j.tig.2007.01.006

Langenberg, A. K., Bink, F. J., Wolff, L., Walter, S., von Wallbrunn, C., Grossmann, M., et al. (2017). Glycolytic functions are conserved in the genome of the wine yeast *Hanseniaspora uvarum*, and pyruvate kinase limits its capacity for alcoholic fermentation. *Appl. Environ. Microbiol.* 83, 1–20. doi: 10.1128/AEM.01580-17

Lee, D. K., Hsiang, T., Lachance, M. A., and Smith, D. R. (2020). The strange mitochondrial genomes of *Metschnikowia* yeasts. *Curr. Biol.* 30, R800–R801. doi: 10.1016/j.cub.2020.05.075

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Ling, J., Daoud, R., Lajoie, M. J., Church, G. M., Soll, D., and Lang, B. F. (2014). Natural reassignment of CUU and CUA sense codons to alanine in *Ashbya* mitochondria. *Nucleic Acids Res.* 42, 499–508. doi: 10.1093/nar/gkt842

Lipinski, K., Kaniak-Golik, A., and Golik, P. (2010). Maintenance and expression of the *S. cerevisiae* mitochondrial genome—from genetics to evolution and systems biology. *Biochim. Biophys. Acta* 1797, 1086–1098. doi: 10.1016/j.bbabio.2009.12.019

Luttik, M. A. H., Overkamp, K. M., Kötter, P., De Vries, S., Van Dijken, J. P., and Pronk, J. T. (1998). The *Saccharomyces cerevisiae NDE1* and *NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *J. Biol. Chem.* 273, 24529–24534. doi: 10.1074/jbc.273.38.24529

Maleszka, R., Skelly, P., and Clark-Walker, G. (1991). Rolling circle replication of DNA in yeast mitochondria. *EMBO J.* 10, 3923–3929. doi: 10.1002/j.1460-2075.1991.tb04962.x

Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* 13, e1002220–e1002226. doi: 10.1371/journal.pbio.1002220

Melvin, R. G., and Ballard, J. W. O. (2017). Cellular and population level processes influence the rate, accumulation and observed frequency of inherited and somatic mtDNA mutations. *Mutagenesis* 32, 323–334. doi: 10.1093/mutage/gex004

Merico, A., Sulo, P., Piškur, J., and Compagno, C. (2007). Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *FEBS J.* 274, 976–989. doi: 10.1111/j.1742-4658.2007.05645.x

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Mossman, J. A., Biancani, L. M., Zhu, C. T., and Rand, D. M. (2016). Mitonuclear epistasis for development time and its modification by diet in *Drosophila*. *Genetics* 203, 463–484. doi: 10.1534/genetics.116.187286

Mukhopadhyay, J., and Hausner, G. (2021). Organellar introns in fungi, algae, and plants. *Cells* 10:2001. doi: 10.3390/cells10082001

Muñoz-Gómez, S. A., Slamovits, C. H., Dacks, J. B., Baier, K. A., Spencer, K. D., and Wideman, J. G. (2015). Ancient homology of the mitochondrial contact site and cristae organizing system points to an endosymbiotic origin of mitochondrial cristae. *Curr. Biol.* 25, 1489–1495. doi: 10.1016/j.cub.2015.04.006

Nguyen, T. H. M., Sondhi, S., Ziesel, A., Paliwal, S., and Fiumera, H. L. (2020). Mitochondrial-nuclear coadaptation revealed through mtDNA replacements in *Saccharomyces cerevisiae*. *BMC Evol. Biol.* 20, 128–112. doi: 10.1186/s12862-020-01685-6

Nguyen, T. H. M., Tinz-Burdick, A., Lenhardt, M., Geertz, M., Ramirez, F., Schwartz, M., et al. (2023). Mapping mitonuclear epistasis using a novel recombinant yeast population. *PLoS Genet.* 19, e1010401–e1010430. doi: 10.1371/journal.pgen.1010401

Nguyen, D. T., Wu, B., Xiao, S., and Hao, W. (2020). Evolution of a record-setting AT-rich genome: Indel mutation, recombination, and substitution bias. *Genome Biol. Evol.* 12, 2344–2354. doi: 10.1093/GBE/EVAA202

Nosek, J., Dinouël, N., Kovac, L., and Fukuhara, H. (1995). Linear mitochondrial DNAs from yeasts: telomeres with large tandem repetitions. *Mol. Gen. Genet.* 247, 61–72. doi: 10.1007/BF00425822

Nosek, J., and Tomáška, L. (2003). Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr. Genet.* 44, 73–84. doi: 10.1007/s00294-003-0426-z

Noutahi, E., Calderon, V., Blanchette, M., Lang, F. B., and El-Mabrouk, N. (2017). CoreTracker: accurate codon reassignment prediction, applied to mitochondrial genomes. *Bioinformatics* 33, 3331–3339. doi: 10.1093/bioinformatics/btx421

Nunn, C. J., and Goyal, S. (2022). Contingency and selection in mitochondrial genome dynamics. *eLife* 11, 1–46. doi: 10.7554/eLife.76557

O'Boyle, S., Bergin, S. A., Hussey, É. E., McLaughlin, A. D., Riddell, L. R., Byrne, K. P., et al. (2018). Draft genome sequence of the yeast *Nadsonia starkeyi-henricii* UCD142, isolated from forest soil in Ireland. *Genome Announc.* 6, 1–2. doi: 10.1128/genomeA.00549-18

Opulente, D. A., LaBella, A. L., Harrison, M.-C., Wolters, J. F., Liu, C., Li, Y., et al. (2023). Genomic and ecological factors shaping specialism and generalism across an entire subphylum. *bioRxiv*:2023.06.19.545611.

Opulente, D. A., Rollinson, E. J., Bernick-Roehr, C., Hulfachor, A. B., Rokas, A., Kurtzman, C. P., et al. (2018). Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biol.* 16:26. doi: 10.1186/s12915-018-0498-3

Paliwal, S., Fiumera, A. C., and Fiumera, H. L. (2014). Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* 198, 1251–1265. doi: 10.1534/genetics.114.168575

Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Peris, D., Arias, A., Orlic, S., Belloch, C., Perez-Traves, L., Querol, A., et al. (2017). Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. *Mol. Phylogenet. Evol.* 108, 49–60. doi: 10.1101/028324

Pfeiffer, T., and Morley, A. (2014). An evolutionary perspective on the Crabtree effect. *Front. Mol. Biosci.* 1, 1–6. doi: 10.3389/fmolb.2014.00017

Pinheiro, J, and Bates, D. (2023). nlme: Linear and Nonlinear Mixed Effects Models.

Pramateftaki, P. V., Kouvelis, V. N., Lanaridis, P., and Typas, M. A. (2006). The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. *FEMS Yeast Res.* 6, 77–90. doi: 10.1111/j.1567-1364.2005.00018.x

Repar, J., and Warnecke, T. (2017). Mobile introns shape the genetic diversity of their host genes. *Genetics* 205, 1641–1648. doi: 10.1534/genetics.116.199059

Roger, A. J., Muñoz-Gómez, S. A., and Kamikawa, R. (2017). The origin and diversification of mitochondria. *Curr. Biol.* 27, R1177–R1192. doi: 10.1016/j.cub.2017.09.015

Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., et al. (2013). Characterizing and measuring bias in sequence data. *Genome Biol.* 14:R51. doi: 10.1186/gb-2013-14-5-r51

Rozpędowska, E., Hellborg, L., Ishchuk, O. P., Orhan, F., Galafassi, S., Merico, A., et al. (2011). Parallel evolution of the make–accumulate–consume strategy in *Saccharomyces* and *Dekkera* yeasts. *Nat. Commun.* 2:302. doi: 10.1038/ncomms1305

Sandor, S., Zhang, Y., and Xu, J. (2018). Fungal mitochondrial genomes and genetic polymorphisms. *Appl. Microbiol. Biotechnol.* 102, 9433–9448. doi: 10.1007/s00253-018-9350-5

Santamaria, M., Lanave, C., Vicario, S., and Saccone, C. (2007). Variability of the mitochondrial genome in mammals at the inter-species/intra-species boundary. *Biol. Chem.* 388, 943–946. doi: 10.1515/BC.2007.121

Scannell, D. R., Zill, O., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., et al. (2011). The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces* sensu stricto genus. *G3 (Bethesda)* 1, 11–25. doi: 10.1534/g3.111.000273

Schikora-Tamarit, M. À., Marcet-Houben, M., Nosek, J., and Gabaldón, T. (2021). Shared evolutionary footprints suggest mitochondrial oxidative damage underlies multiple complex i losses in fungi. *Open Biol.* 11:200362. doi: 10.1098/rsob.200362

Sharp, N. P., Sandell, L., James, C. G., and Otto, S. P. (2018). The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5046–E5055. doi: 10.1073/pnas.1801040115

Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., et al. (2018). Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 175, 1533–1545.e20. doi: 10.1016/J.CELL.2018.10.023

Shulgina, Y., and Eddy, S. R. (2021). A computational screen for alternative genetic codes in over 250,000 genomes. *eLife* 10, 1–25. doi: 10.7554/eLife.71402

Solieri, L. (2010). Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends Microbiol.* 18, 521–530. doi: 10.1016/j.tim.2010.08.001

Steenwyk, J. L., Opulente, D. A., Kominek, J., Shen, X.-X., Zhou, X., Labella, A. L., et al. (2019). Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol.* 17:e3000255. doi: 10.1371/journal.pbio.3000255

Strope, P. K., Skelly, D., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., et al. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774. doi: 10.1101/gr.185538.114

Szabóová, D., Hapala, I., and Sulo, P. (2018). The complete mitochondrial DNA sequence from *Kazachstania sinensis* reveals a general +1C frameshift mechanism in CTGY codons. *FEMS Yeast Res.* 18, 1–10. doi: 10.1093/femsyr/foy028

Valach, M., Farkas, Z., Fricova, D., Kovac, J., Brejova, B., Vinar, T., et al. (2011). Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Res.* 39, 4202–4219. doi: 10.1093/nar/gkq1345

van de Vossenberg, B. T. L. H., Brankovics, B., Nguyen, H. D. T., van Gent-Pelzer, M. P. E., Smith, D., Dadej, K., et al. (2018). The linear mitochondrial genome of the quarantine chytrid *Synchytrium endobioticum*; insights into the evolution and recent history of an obligate biotrophic plant pathogen. *BMC Evol. Biol.* 18:136. doi: 10.1186/s12862-018-1246-6

Visinoni, F., and Delneri, D. (2022). Mitonuclear interplay in yeast: from speciation to phenotypic adaptation. *Curr. Opin. Genet. Dev.* 76:101957. doi: 10.1016/j.gde.2022.101957

Wick, R. R., and Holt, K. E. (2019). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res.* 8, 1–22. doi: 10.12688/f1000research.21782.1

Wolfe, K. H. (2015). Origin of the yeast whole-genome duplication. *PLoS Biol.* 13, e1002221–e1002227. doi: 10.1371/journal.pbio.1002221

Wolters, J. F., Chiu, K., and Fiumera, H. L. (2015). Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genomics* 16:451. doi: 10.1186/s12864-015-1664-4

Wu, B., Buljic, A., and Hao, W. (2015). Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. *Mol. Biol. Evol.* 32, 2559–2570. doi: 10.1093/molbev/msv127

Wu, B., and Hao, W. (2014). Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. *G3 (Bethesda)* 4, 605–612. doi: 10.1534/g3.113.009910

Wu, B., and Hao, W. (2019). Mitochondrial-encoded endonucleases drive recombination of protein-coding genes in yeast. *Environ. Microbiol.* 21, 4233–4240. doi: 10.1111/1462-2920.14783

Xiao, S., Nguyen, D. T., Wu, B., and Hao, W. (2017). Genetic drift and indel mutation in the evolution of yeast mitochondrial genome size. *Genome Biol. Evol.* 9, 3088–3099. doi: 10.1093/gbe/evx232

Xue, L., Moreira, J. D., Smith, K. K., and Fetterman, J. L. (2023). The mighty NUMT: mitochondrial DNA flexing its code in the nuclear genome. *Biomol. Ther.* 13, 1–11. doi: 10.3390/biom13050753

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Zachar, I., and Szathmáry, E. (2017). Breath-giving cooperation: critical review of origin of mitochondria hypotheses. *Biol. Direct* 12:19. doi: 10.1186/s13062-017-0190-5