


# Mitogenomics and mitochondrial gene phylogeny decipher the evolution of *Saccharomycotina* yeasts

Anastasia C. Christinaki<sup>1,†</sup>, Spyros G. Kanellopoulos<sup>1,†</sup>, Alexandra M. Kortsinoglou<sup>1,†</sup>, Marios A. Andrikopoulos<sup>1</sup>, Bart Theelen<sup>2</sup>, Teun Boekhout<sup>2,3</sup>, and Vassili N. Kouvelis <sup>1,\*</sup>

<sup>1</sup>Faculty of Biology, Department of Genetics and Biotechnology, National and Kapodistrian University of Athens, Athens, Greece

<sup>2</sup>Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands

<sup>3</sup>Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands

<sup>†</sup>A.C.C., S.G.K. and A.M.K. have contributed equally to this work.

\*Corresponding author: E-mail: kouvelis@biol.uoa.gr.

Accepted: 12 May 2022

## Abstract

*Saccharomycotina* yeasts belong to diverse clades within the kingdom of fungi and are important to human everyday life. This work investigates the evolutionary relationships among these yeasts from a mitochondrial (mt) genomic perspective. A comparative study of 155 yeast mt genomes representing all major phylogenetic lineages of *Saccharomycotina* was performed, including genome size and content variability, intron and intergenic regions' diversity, genetic code alterations, and syntenic variation. Findings from this study suggest that mt genome size diversity is the result of a ceaseless random process, mainly based on genetic recombination and intron mobility. Gene order analysis revealed conserved syntenic units and many occurring rearrangements, which can be correlated with major evolutionary events as shown by the phylogenetic analysis of the concatenated mt protein matrix. For the first time, molecular dating indicated a slower mt genome divergence rate in the early stages of yeast evolution, in contrast with a faster rate in the late evolutionary stages, compared to their nuclear time divergence. Genetic code reassignments of mt genomes are a perpetual process happening in many different parallel evolutionary steps throughout the evolution of *Saccharomycotina*. Overall, this work shows that phylogenetic studies based on the mt genome of yeasts highlight major evolutionary events.

**Key words:** mitochondrial genomes, fungi, molecular dating, phylogenomics, synteny, codon usage.

## Significance

Mitochondria (mt) are organelles of eukaryotic cells that play an important role in the efficient production of ATP. Yeasts are conditional aerobes, but functional mt are of utmost importance. Mt originated from an ancestral  $\alpha$ -proteobacterial endosymbiont and they are semi-autonomous, since they carry their own mt genomes. The mitogenomes of yeasts provide insights to the still incomplete picture of their phylogeny and evolution. In this study, a phylogenetic analysis of mt genes, a comparative genome analysis concerning gene content and order, intron abundance and gene codon usage, combined with molecular dating helped deciphering the evolution of mt in yeasts. This work showed that mt genomes evolved in a perpetual manner with different divergence rates compared to nuclear genomes.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Introduction

Ascomycetous yeasts comprise some of the most economically and clinically important species within the fungal kingdom, which influence many aspects of humans' everyday life. For instance, the production of some of the most common foods in human diet, like bread and wine, involves *Saccharomyces cerevisiae* (Botstein and Fink 2011; Peter et al. 2018), while other yeasts contribute to the fermentation and flavoring of alcoholic beverages (Eldarov et al. 2016). Other members of the *Saccharomycotina* subphylum, like some *Candida* species, cause serious human infections (Wang et al. 2018; Pote et al. 2020). Besides their importance to humans, *Saccharomycotina* yeasts present several other characteristics like (usually) a single cell morphology, asexual and sexual reproduction by budding and the formation of ascospores, respectively, and facultative respiration. These characteristics render them highly interesting for research, since they can be employed as single-cell eukaryotic models, with *S. cerevisiae* as the most commonly used example (Stajich et al. 2009; Kurtzman and Boekhout 2017; Nielsen 2019).

According to current taxonomy, *Saccharomycotina* consists of a single order, Saccharomycetales, in contrast with other fungal subphyla that contain several discernible orders. The main reason for this taxonomy is the "simple" morphology of the budding yeast cells, lacking sufficient morphological discrimination (Knop 2011). Molecular based phylogenies classify yeasts at the family level. Recent phylogenetic studies divide this subphylum into 12 families (Shen et al. 2018), in contrast to previous studies which were based on few nuclear molecular markers that assigned them into 11 different families (Kurtzman 2011; Kurtzman and Boekhout 2017). Numerous studies during the last two decades showed that *Saccharomycotina* presents high levels of genetic variability (Benítez et al. 1996; Braun 2003; Teresa Fernández-Espinar et al. 2003; Dujon 2006). Contrasting morphological similarity between species like *Yarrowia lipolytica* and *S. cerevisiae*, however their genetic differentiation has reached extensive levels, similar to that between roundworm and human genome (Dujon 2006; Shen et al. 2018). This genetic diversity is attributed to several evolutionary events that took place within this lineage, like the alteration of their genetic code and whole genome duplication (WGD) event (Kellis et al. 2004; Wolfe 2004; Krassowski et al. 2018). In detail, genetic code alterations occurred in three independent evolutionary events, during which the CUG codon translated Serine (two different events) or Alanine (third event), instead of Leucine (universal genetic code). As a result, phylogenetic trees support the existence of the CUG-Ser1, CUG-Ser2, and CUG-Ala distinct groups (Santos et al. 1997; Miranda et al. 2006; Krassowski et al. 2018). The second evolutionary landmark which influenced genetic

diversity in yeasts is the WGD event that happened ~100 million years ago (MYA) in Saccharomycetaceae (Wolfe and Shields 1997; Kellis et al. 2004). It is proposed that WGD is the result of autopolyploidization or hybridization. Recently, a model was proposed to explain this WGD (Marcet-Houben et al. 2015), with genetic rearrangements, gene losses, and mutations following a hybridization event that led to genome evolution and adaptation (Wolfe 2015; Fisher et al. 2018; Escalera-Fanjul et al. 2019).

The mitochondrion (mt) is a vital organelle due to its role in energy production of eukaryotic cells (Hatefi 1985). Multiple mt exist inside a cell containing their own genome (mtDNA) in multiple replicates, with genes encoding, among other products, proteins necessary for oxidative phosphorylation (de Zamaroczy and Bernardi 1985; Burger et al. 2003). The first sequenced mtDNAs of the *Saccharomycotina* subphylum were those of *Wickerhamomyces canadensis* and *S. cerevisiae* (Sekito et al. 1995; Foury et al. 1998). The mt genes located in the genomes of *Saccharomycotina* species are the genes typically found among the majority of fungal and metazoan lineages, i.e., subunits of the adenosine triphosphate (ATP) synthase complex (*atp6*, *atp8*, and *atp9*), apocytochrome b (*cob*), cytochrome c oxidase complex (*cox1-3*), nicotinamide adenine dinucleotide (NADH) dehydrogenase (*nad1-6*, *nad4L*), the small and large rRNA (*rns* and *ml*), and approximately 25 tRNAs (Kouvelis et al. 2004; Dujon 2010; Solieri 2010; Korovesi et al. 2018). Interestingly, mt genes encoding subunits of NADH dehydrogenase are absent in specific *Saccharomycotina* mt genomes, i.e., species belonging to families Saccharomycetaceae and Saccharomycodaceae (Pramateftaki et al. 2006; Freel et al. 2015).

The mtDNA has been utilized for comparative genomics of the phylum Ascomycota at various taxonomic levels (Freel et al. 2015; Wolters et al. 2015; Hao 2022), as it has a different evolution rate compared to the nuclear genome, which makes mtDNA a valuable alternative source of phylogenetic markers as previously shown (Ballard and Whitlock 2004; Kouvelis et al. 2008; Duò et al. 2012; Kortsinoglou et al. 2019). The size of the mtDNA varies greatly among *Saccharomycotina* species, i.e., from 18,884 bp (*Hanseniaspora uvarum*) to 107,123 bp (*Nakaseomyces bacillisporus*) up to now (Pramateftaki et al. 2006; Bouchier et al. 2009). The majority of yeast mt genomes are reported as circular molecules (Williamson 2002). *Candida parapsilosis* (Kováč et al. 1984) and *Hansenula mrakii* (Wesolowski and Fukuhara 1981) were the first two described yeast species presenting linear mtDNAs. Recent studies suggest that mt genomes of yeast are linear, or circular mapping, or interconvert between the two topologies (Nosek et al. 1998; Nosek and Tomáška 2003; Shen et al. 2018). Moreover, the mt gene order (synteny) is diverse among taxa, even within the

same genus (Fricova et al. 2010; Bartelli et al. 2013; Freel et al. 2015). It has been observed that certain gene clusters are formed, that remain conserved despite the general occurring relocations of mt regions (Pantou et al. 2008). These existing patterns help to identify gene rearrangements and partial genome inversions and duplications. The above characteristics, as well as the frequent intra- and inter-species recombination of mtDNA (Fritsch et al. 2014; Brankovics et al. 2017), render mtDNA helpful for resolving phylogenetic relationships of *Saccharomycotina* yeasts.

This work aims to study the evolutionary relationships of ascomycetous yeasts (in the rest of the manuscript referred to as “yeasts”) from an mt genomic perspective. The questions to be answered concern the evolution of mt genomes and their features, as well as their potential contribution to the evolution of *Saccharomycotina*. The resolution of the phylogenetic positioning of all families, as supported by the concatenated mt gene matrix, is also addressed. This will be combined with the timeframe during which the evolutionary changes have happened, as shown by the mt genes used in the phylogenetic matrix. To this end, this study investigates the genetic elements (i.e., introns, gene clusters, intergenic regions, GC islands), genetic mechanisms (i.e., recombination) and evolutionary events (i.e., genetic code changes) under the prism of molecular dating. Moreover, it examines their impact on shaping the existing phylogenetic relationships within this subphylum.

## Results

In order to investigate the evolution of *Saccharomycotina* subphylum from a mt perspective, mt genomes of 155 yeast species and strains were examined, representing all known phylogenetic families. The genome sizes varied greatly with the smallest and largest being the ones of *Hanseniaspora pseudoguilliermodii* (11,223 bp—whole genome sequencing [WGS] from Shen et al. 2018; mt retrieval and annotation in this work, see doi: [10.6084/m9.figshare.19714219](https://doi.org/10.6084/m9.figshare.19714219)) and *Saccharomycopsis malanga* (173,850 bp—CP025327; annotation in this work, see doi: [10.6084/m9.figshare.19714219](https://doi.org/10.6084/m9.figshare.19714219)), respectively. The size of roughly half of the genomes examined (53.8%) was distributed between 20–39.9 kb. When including sizes ranging from 40 to 40.9 kb, 70.7% of all assessed genomes fit this size distribution ([supplementary table S1, Supplementary Material](#) online). Genome diversity was further analyzed along with synteny, introns intergenic regions, phylogeny and molecular dating. Furthermore, 20 species representing the other Ascomycota subphyla, i.e., 10 Pezizomycotina and 10 Taphrinomycotina, were used as outgroups.

### Mt concatenated gene phylogeny

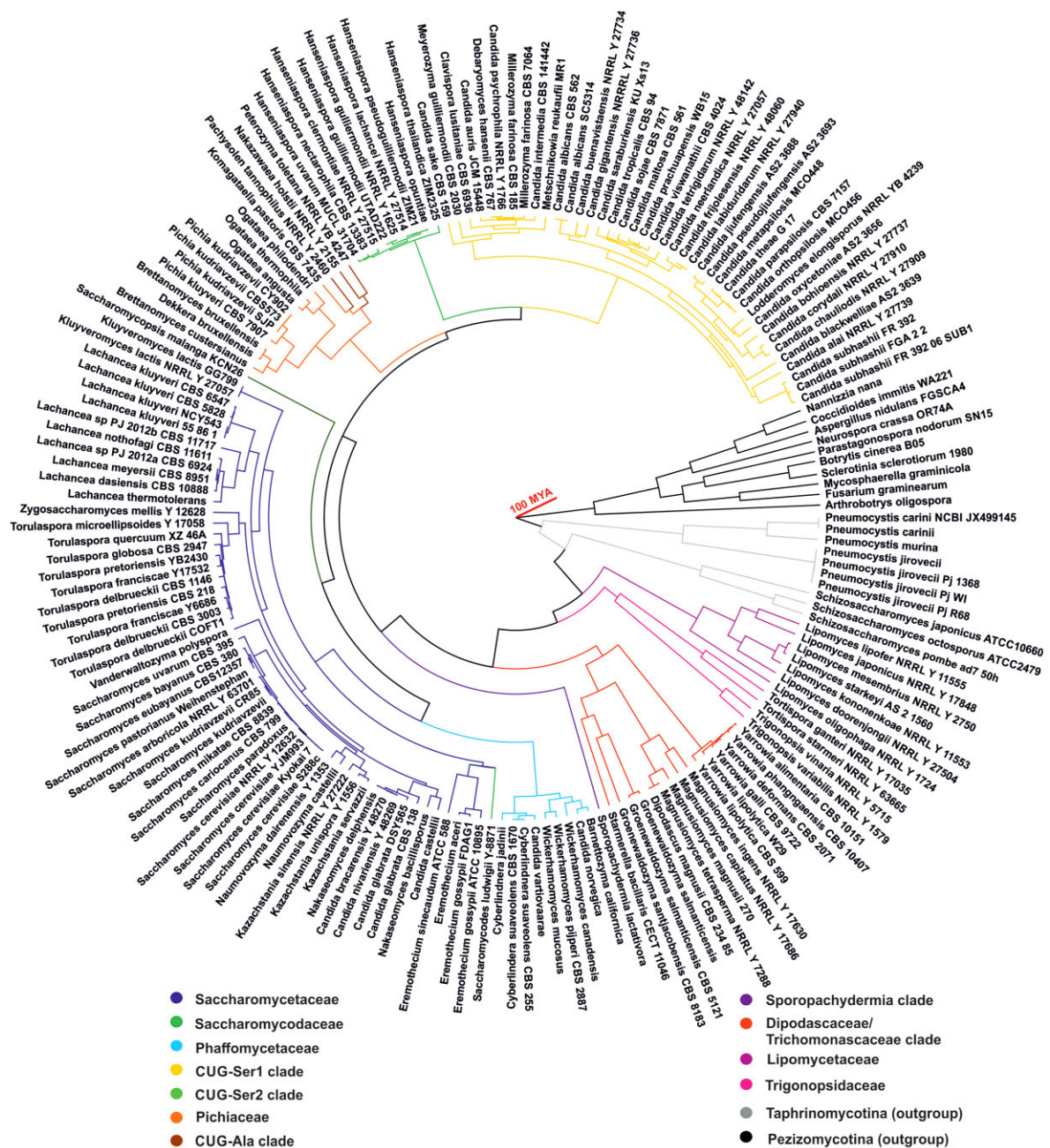
The sequences of the 14 conserved mt proteins were used to determine the phylogenetic relationships of yeasts. Single

gene phylogenies presented the phylogenetic signal of each gene itself, and in order to represent the phylogeny of the whole mt genome, a concatenated approach was followed. The tree produced by Bayesian analysis was well supported, as almost all clade topologies presented posterior probability values of 100% (Fig. 1). This result was mostly in accordance with the trees produced by the Neighbor-Joining (NJ) and Maximum Likelihood (ML) methods. In the Bayesian-based tree, the 155 yeast taxa formed 11 phylogenetic clades with almost identical topologies to those obtained by the nuclear based phylogeny (Shen et al. 2018), with only the three following differences: 1) the *Hanseniaspora* clade was placed basally to the CUG-Ser1 clade apart of *Saccharomycodes ludwigii*, the other member of the family, 2) *Komagataella pastoris* appeared to be inside the CUG-Ala clade and not in its closely related family (Pichiaceae), and 3) the KLE (*Lachancea*, *Kluyveromyces*, and *Eremothecium*) Saccharomycetaceae group is not monophyletic, but the *Kluyveromyces* clade is basal to the *Lachancea* clade and *Eremothecium* spp. are placed in a sister clade of *S. ludwigii* (Fig. 1). Based on the only available mt genome of *S. malanga*, mt based phylogeny confirmed that the CUG-Ser2 clade was basal to the clade containing the families Saccharomycetaceae and Phaffomycetaceae and it did not seem to be related to the CUG-Ser1 clade, as was also supported by the nuclear data (Shen et al. 2018).

### Molecular dating

Molecular dating was used to estimate the divergence of the *Saccharomycotina* subphylum, presented by the mt phylogeny and calculated by the Rel-Time method of MEGA11 and the LSD2 method in IQ-TREE (Figs. 1 and 2 and [supplementary file S2 and table S2, Supplementary Material](#) online). For estimating the molecular clock of this tree, the same calibration point used in the work of Shen et al. (2018) were employed. For simplicity reasons, in the following text the divergence times will be presented as RelTime/LSD2 based divergence times. Main findings worth mentioning were 1) the subphylum originated approximately 567/471 MYA, 2) the WGD group diverged 74/83 MYA from the ZT clade (*Zygosaccharomyces/Torulasporea*) of the Saccharomycetaceae, 3) the CUG-Ser1 and CUG-Ser2 clades diverged independently from their closest relatives (Pichiaceae and Saccharomycetaceae/Phaffomycetaceae) at 227/241 and 213/241 MYA, respectively and, 4) the CUG-Ala clade diverged from Pichiaceae at 137/174 MYA (Fig. 2). In brief, the outcome of all analyses of molecular dating showed that mt phylogenetic dating is significantly different from the nuclear one and the mt phylogeny revealed a faster and slower evolutionary rate in later and earlier divergence events. Mt molecular dating using the RelTime method seemed to be closer to the nuclear dating than when using the LSD2 method, possibly due to the





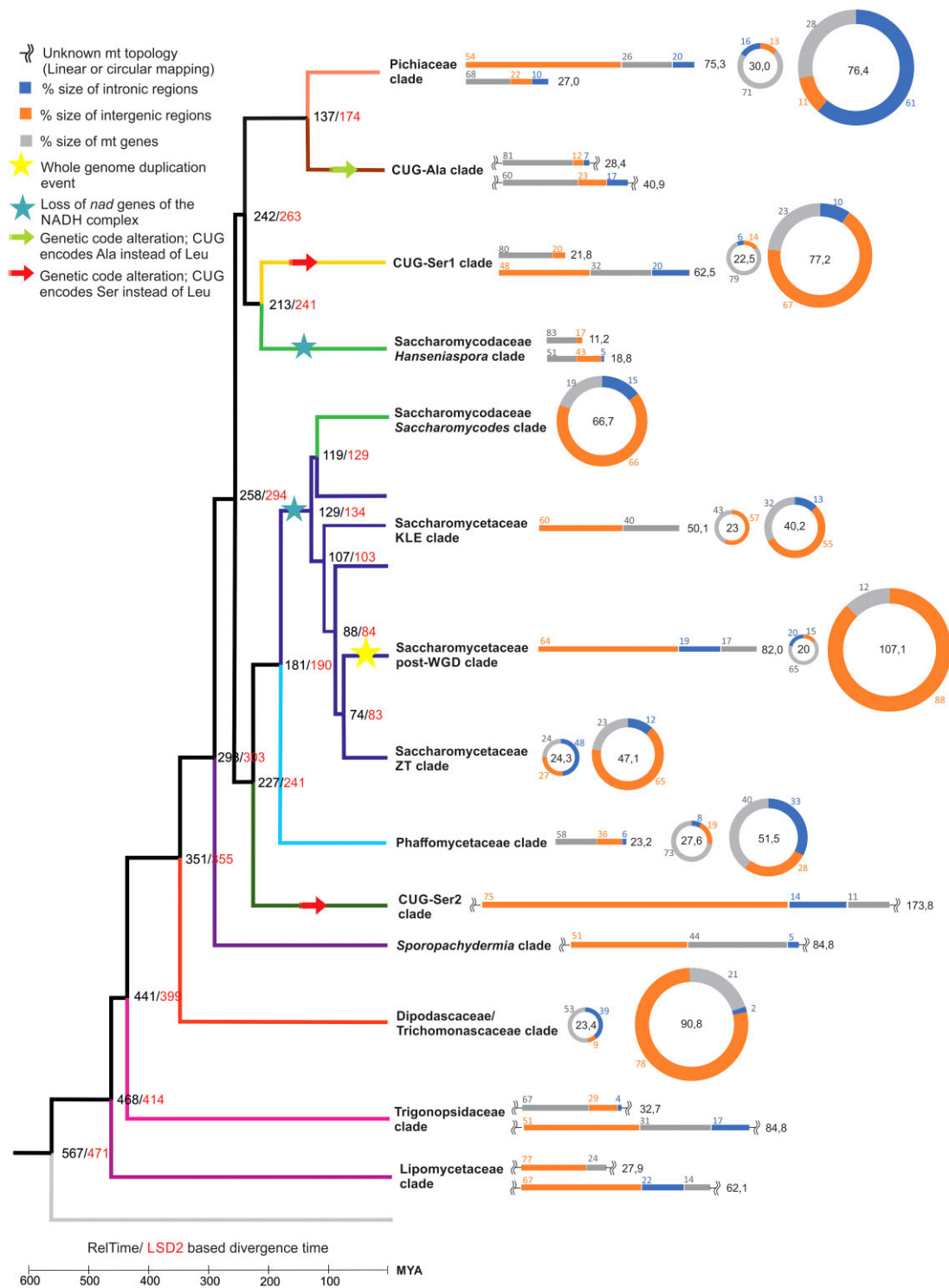
**Fig 1.**—Time-calibrated phylogenetic tree of all yeast species based on the concatenated dataset of the 14 conserved mt proteins (i.e., Atp6,8–9, Cob, Cox1–3, Nad1–6 and 4L) as produced by the Bayesian inference (BI) method. Branch lengths represent the divergence time of each node. All topologies produced are in accordance with the respective NJ and ML methods. The species belonging to Taphrinomycotina and Pezizomycotina subphyla were used to root the tree. The clades are colored according to taxonomic families as follows: Saccharomycetaceae blue, Saccharomycodaceae dark green, Phaffomycetaceae light blue, CUG-Ser1 clade yellow, CUG-Ser2 clade light green, Pichiaceae orange, Sporopachydermia clade dark purple, Dipodascaceae/Trichomonascaceae clade red, Lipomycetaceae light purple, Trigonopsidaceae pink, CUG-Ala clade brown and Taphrinomycotina/Peizomycotina gray (outgroup).

different algorithms used by these two methods (Fig. 1 and supplementary file S1, Supplementary Material online).

### General characteristics of mt genomes

The 155 yeast mt genomes studied in this work presented a few similar characteristics and several different

properties. In all fungal mt genomes, the content in protein coding genes related to oxidative phosphorylation and production of ATP was conserved, with the single exception of the genes coding subunits of the NADH complex which were absent in the families Saccharomycetaceae and Saccharomycodaceae. Based



**FIG 2.**—Mt genome variability of each *Saccharomycotina* phylogenetic group as shown with phylogeny. The percentage of sizes of intergenic, intronic and gene regions are presented in the linear or circular mapped mt genomes. The largest and smallest mt genomes of each cluster are selected in order to indicate the range of sizes and mt genome topology of each phylogenetic group. Node values correspond to the divergence time of each node (RelTime based and LSD2 based values are provided with black and red color, respectively). The main evolutionary events in *Saccharomycotina* subphylum (i.e., WGD, genetic code alterations and *nad* gene loss) are mapped on the phylogenetic tree. The number inside or next to the mt genomes refer to mt genome sizes in kb. The time calibration bar is provided at the bottom of the figure.

on the mt molecular dating, this gene loss was dated 181/190 MYA (Fig. 2).

Most yeast families presented a uniform pattern of presence or absence of *rps3*, a gene that contributes to the assembly of the small ribosomal subunit. Specifically, this gene was present in members of the families Phaffomycetaceae, Dipodascaceae, Pichiaceae, Saccharomycetaceae and Trichomonascaceae, while it was absent in all other yeast families. One exception was found in the Saccharomycodaceae family represented here by two genera, i.e., *Saccharomycodes* and *Hanseniaspora*, in which the gene *rps3* was present in the first genus and absent in the latter. The number of *trn* genes in all examined species fluctuated between 21 and 31, with an average of 24 (supplementary table S1, Supplementary Material online).

Gene content of yeast mt genomes was mostly conserved (with the exception of Saccharomycetaceae and Saccharomycodaceae, which lack the *nad* genes). If introns and intergenic regions were not considered, all mt genomes were approximately of the same size. The mean size of all 14 mt protein coding genes combined was approximately 16.5 kb. Therefore, the observed genome size variability was determined by the diversity of introns and intergenic regions. For instance, the examined mt genome of *N. bacillisporus* (107,123 kb) presented intergenic regions equaling 88% of the entire genome, and no intronic sequences. The genome with the smallest proportion of intergenic regions was *Millerozyma farinosa*, with only 7% attributed to intergenic regions, while 48% of its genome size corresponded to introns (Fig. 2 and supplementary table S1, Supplementary Material online).

Yeast genomes were mapped as either linear or circular (19 and 112 mt genomes, respectively, the topology of the remaining 24 is not yet known) (Fig. 2 and supplementary table S1, Supplementary Material online). It appeared that both size and genome structure variation occur scattered among all yeast families without a clear pattern. However, the general size tendency of mt genomes was clearly towards 20–50 kb, and the majority of them were of circular configuration (supplementary table S1, Supplementary Material online).

### Synteny

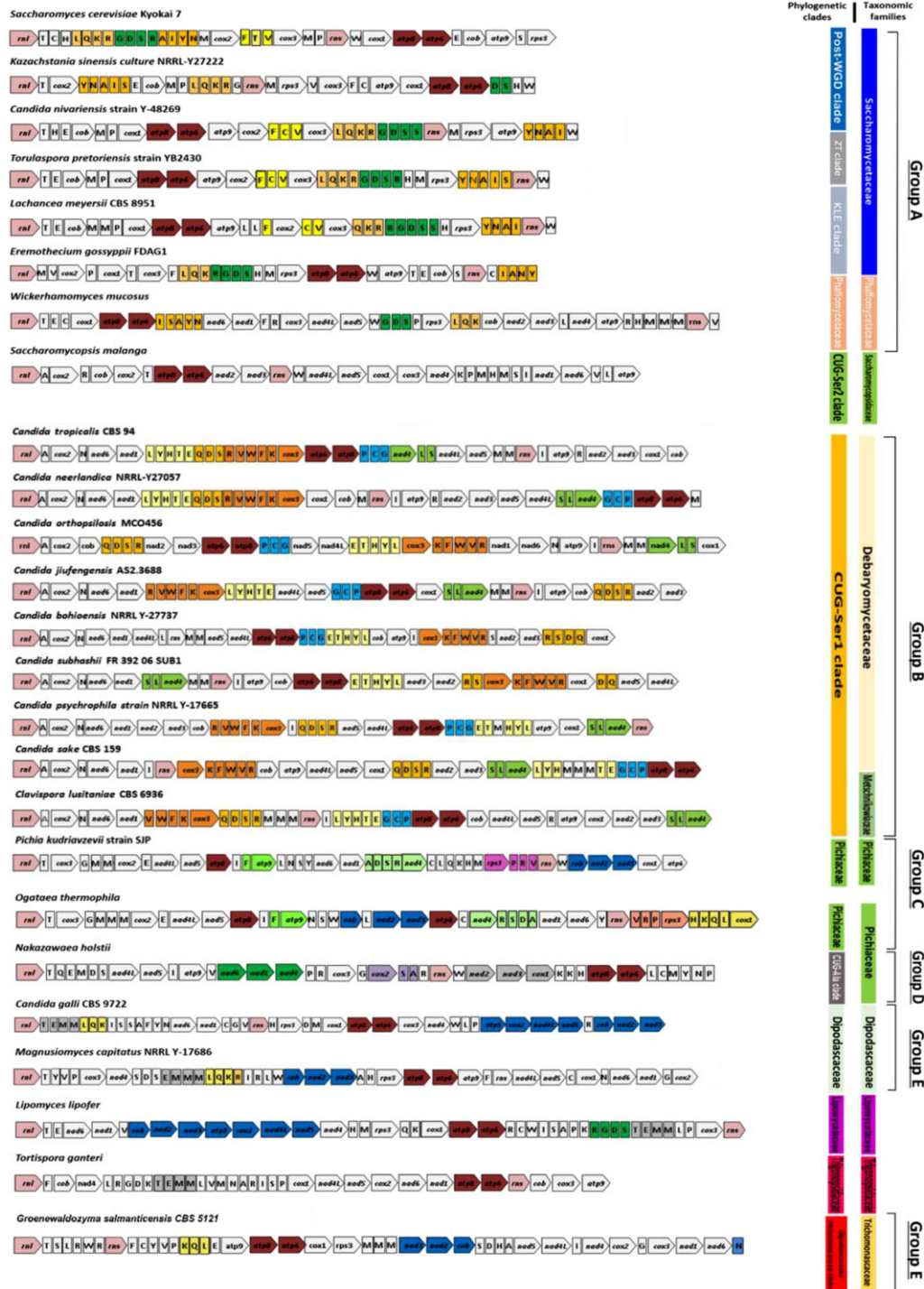
Gene order analysis revealed a clear pattern of genome organization among all taxa studied. This pattern was based on the existence of several gene clusters consisting of both *trn* genes, protein and ribosomal coding genes. These genes were considered to form clusters, because they were found intact (at the same or inverted orientation) in different locations of mt genomes of various species. Based on the presence of these conserved clusters, six major groups and several sub-groups of taxa could be

determined (Fig. 3, Table 1 and supplementary table S3, Supplementary Material online). The only criterion for grouping various taxa in one main group was the existence of at least two identical or highly similar syntenic gene groups in all its members (Table 1 and supplementary table S3, Supplementary Material online). Therefore, Group A contained species belonging to clades Saccharomycetaceae post-WGD, ZT, KLE, and Phaffomycetaceae, and Group B consisted of the Metschnikowiaceae and Debaryomycetaceae families. Group C corresponded to species of Pichiaceae, while species that belong to the CUG-Ala clade comprised Group D. Finally, species belonging to the Dipodascaceae and Trichomonascaceae families represented group E, and the Lipomycetaceae clade represented Group F (Fig. 3 and supplementary table S3, Supplementary Material online). Species belonging to CUG-Ser 2, Sporopachydermia and Trigonopsidaceae clades were underrepresented and did not present enough similarities to assign them to any of the above-mentioned groups. For this reason, it seemed prudent that these groups may be considered as separate additional groups which, however, need further examination by adding sequence data from more species (not currently publicly available). It is worth mentioning that *Hanseniaspora* spp., which belong to Saccharomycodaceae clade, were not considered as an additional syntenic group, despite being well represented (all species have publicly available mt genome data, Nguyen et al. 2020). The main reason for this is that all *Hanseniaspora* spp. have an almost identical gene order and did not present any evident genomic rearrangements that would allow the recognition of separate gene clusters similar to the syntenic gene shuffling found in other relative clades, like the Debaromycetaceae clade which forms Group B.

Taxa belonging to Group A had a small degree of intra-species rearrangements in all subgroups (Fig. 3 and supplementary table S3, Supplementary Material online). It was possible to distinguish four main gene clusters (Table 1). Despite the conserved anchoring of these clusters at intra species level, there was a high rate of gene rearrangements inside each individual cluster. The gene content was the same, but the gene order was different in each syntenic group. Characteristic examples were the rearrangements of syntenic groups *trnG-trnD-trnS-trnR* and *trnY-trnN-trnA-trnI* (see Table 1 and supplementary table S3, Supplementary Material online for alternative orders).

As previously mentioned, taxa belonging to Group B had seven main syntenic groups (Fig. 3, Table 1 and supplementary table S3, Supplementary Material online). These maintained a conserved gene content but were subjected into several major whole cluster rearrangements. As a result, this gene shuffling created new reversed, joined or divided clusters. Taxa belonging to groups C, D, E and F





**FIG 3.**—Synteny of mt genomes for representative species of all families/phylogenetic clades. Genes are shown as boxed arrows and conserved synteny units are highlighted with different colors. The criterion for grouping various taxa in one group was the existence of at least two identical or highly similar synteny gene groups.

presented larger synteny units of usually four to eight genes (Fig. 3, Table 1 and supplementary table S3, Supplementary Material online). The cluster of *trnR-trnG-trnD-trnS* is

conserved among members of Group F (Lipomycetaceae clade) and Group A (Saccharomycetaceae post WGD, ZT, and KLE clades).

**Table 1**  
Groups of taxa (A–H) and the syntenic gene clusters (1–7) formed by comparing the genome organization of species examined in this study

	Group A	Group B	Group C	Group D	Group E	Group H
1	<i>trnF</i> -( <i>trnC</i> )-(trnV) <sup>a</sup>	<i>atp6-atp8-trnP-trnC-trnG</i>	<i>rps3-trnP-trnR-trnV</i>	<i>nad6-nad1-nad4</i>	<i>cob-nad2-nad3</i>	<i>trnT-trnE-trnM-trnM</i>
2	( <i>trnL</i> )- <i>trnQ-trnK</i> -( <i>trnR</i> )	<i>trnL-trnY-trnH</i> -(trnM)- <i>trnT-trnE</i>	( <i>cob</i> )- <i>nad2-nad3</i>	<i>cox2-trnS-trnA</i>	<i>atp9-cox2-nad4L-nad5<sup>c</sup></i>	<i>trnR-trnG-trnD-trnS</i>
3	<i>trnG-trnD-trnS</i>	• ( <i>trnR</i> )- <i>trnV-trnW-trnF</i> -( <i>trnK-cox3</i> ) • ( <i>trnR</i> )- <i>trnV-trnW-trnF</i> -( <i>trnR</i> )	( <i>nad4</i> )-(trnR)- <i>trnS-trnD-trnA</i>	<i>nad2-nad3-cox1</i>	( <i>trnT</i> )- <i>trnE-trnM-trnM-trnL-trnQ-trnK</i>	<i>cob-nad2-nad3-atp9-cox2-nad4L-nad5-nad4</i>
4	<i>trnY-trnN-trnA-trnI</i> -(trnS) • <i>trnA-trnI-trnY-trnN</i> • <i>trnI-trnS-trnA-trnY-trnN</i> • <i>trnI-trnA-trnN-trnY</i>	<i>trnQ-trnD-trnS</i> -(trnR)	(trnL)- <i>trnQ-trnK-trnH</i> -(trnM)			<i>trnW-trnI-trnS-trnA-trnF-trnL</i>
5		<i>trnS-trnL-nad4</i>				
6		<i>cox2-trnN-nad6-nad1</i>				
7		• ( <i>trnM</i> )- <i>rns-trnI</i> <sup>b</sup> • <i>trnM<sub>1</sub>-trnM<sub>2</sub>-rns</i> • <i>trnM<sub>1</sub>-trnM<sub>2</sub>-rns-trnI</i>				

<sup>a</sup>Except for *Kluyveromyces* spp. and *Eremothecium* spp. and Phaffomycetaceae family.

<sup>b</sup>Except for *C. buenavistae* NRRL Y-27734, *C. albicans* SCS5314 and CBS 562.

<sup>c</sup>Only in *C. galli* and *Yarrowia* spp.



### Genetic codes and codon frequency

The study of 155 strains belonging to the *Saccharomycotina* subphylum showed that the species belonging to families Debaryomycetaceae, Metschnikowiaceae (i.e., both comprising phylogenetic clade CUG-Ser1), Dipodascaceae, Lipomycetaceae, Pichiaceae, Phaffomycetaceae, Trichomonascaceae, Trigonopsidaceae, and the CUG-Ala phylogenetic clade employ genetic code NCBI 4 for translating their mt genes, while almost all species of Saccharomycetaceae (i.e., KLE, ZT and post-WGD clades with the exception of *Kluyveromyces* spp), Saccharomycodaceae (i.e., *Saccharomycodes ludwigii*, *Hanseniaspora* spp.) and *Pachysolen tannophilus* use genetic code NCBI 3 (supplementary tables S1 and S4, Supplementary Material online).

Accumulative relative synonymous codon usage (RSCU) values of individual genes, as well as the overall RSCU values, presented highly biased frequencies for ATG (Met), TGA (Trp), ACA, and ACT (Thr), with the exception of genes *atp8* and *atp9* in the TGA codon, where no codons or Trp were detected (Fig. 4 and supplementary table S4, Supplementary Material online). The ATA codon was used by species utilizing genetic code NCBI 3 with an approximate ratio of 1:5 to the synonymous ATG codon. In a similar way, the TGG codon, which is the universal codon for tryptophan, was mainly found in yeast genomes utilizing genetic code NCBI 4, in a ratio of 1:4 to the respective TGA codon. On the contrary, species using genetic code NCBI 3 had a higher percentage of the TGA codon (97.6%), compared to the TGG (2.3%) (Fig. 4). In all taxa studied, the most abundant Thr codons were ACA and ACT (Fig 4). As expected, CTN codons (Thr) were only detected in species using genetic code NCBI 3, and the most abundant codons, after ACA and ACT, were CTA and CTT (Fig. 4 and supplementary table S4, Supplementary Material online). Trp codons were not detected in *atp9*, while in *atp8* these codons were absent only in species utilizing genetic code NCBI 4 (supplementary table S4, Supplementary Material online). The gene *atp9* had a high number of ACA codons in taxa belonging to both genetic codes.

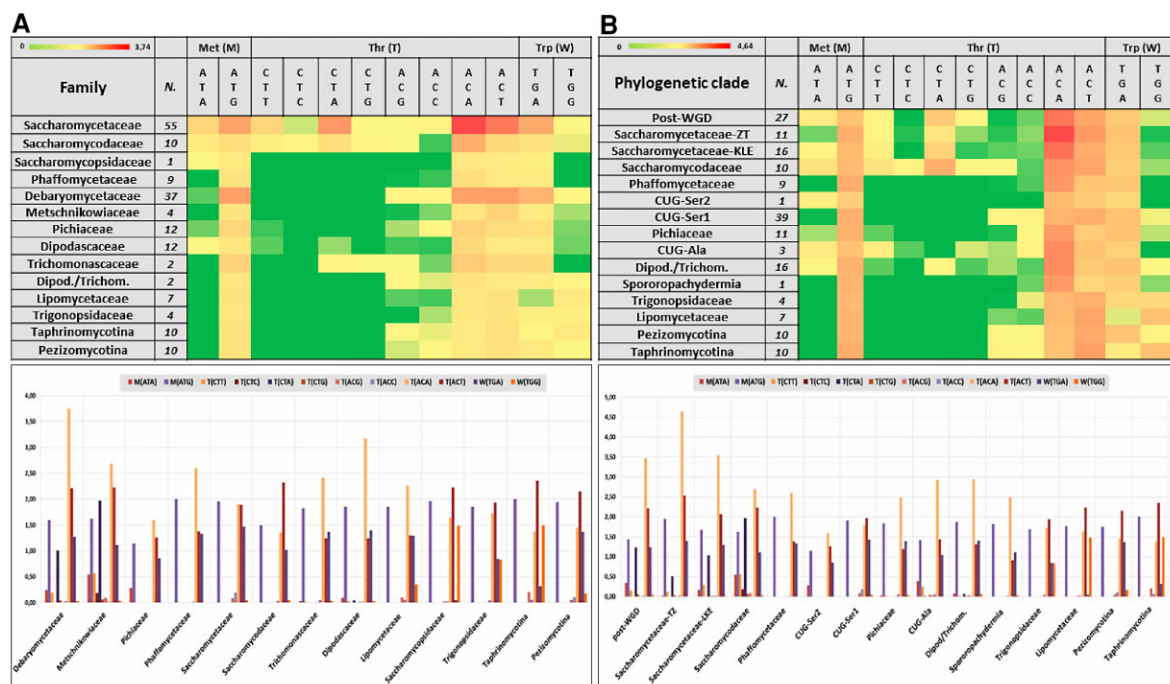
### Intron distribution

The mt genes of the examined yeasts presented an interesting intronic distribution concerning intron types and their host genes. Most introns were found in the *cox1* gene (544 introns) which usually contained IB introns (408 introns) and to a lesser extent type II and ID introns (i.e., 60 and 50, respectively) (supplementary table S5I.b, Supplementary Material online). The *cob* gene was the second most intron-rich gene (200 introns), followed by *rnl* (72 introns) and *nad5* (37 introns) (supplementary table S5I.b, Supplementary Material online). Genes *atp6*, *cox2*, *cox3*, *nad1*, *nad4*, and *rns* contained fewer introns (i.e.,

less than 10) compared to the above genes, while *atp8*, *atp9*, *nad2*, *nad3*, *nad6*, and *nad4L* lacked introns in the mt genomes of the yeasts examined (supplementary table S5I.b, Supplementary Material online). The most common type of intron in the *cob* gene was ID (85 ID out of 200 introns found in *cob*), followed by the IB intron (59) (supplementary table S5I.b, Supplementary Material online). The most abundant intron of *rnl* was IA (50 out of 72 introns found in *rnl*) (supplementary table S5I.b, Supplementary Material online). Eighty-eight group II introns were found in six different genes, i.e., *cox1* (60 introns), *cob* (17 introns), *rnl* (8 introns) and *rns*, *cox2* and *nad5* (from 1 intron each) (supplementary table S5I.b, Supplementary Material online).

IB introns were the most abundant within various mt genes of all genomes studied, regardless of the family or phylogenetic clade (supplementary table S5I.a, Supplementary Material online). In detail, in all families IB introns represented more than 50% of the total introns found, with the lowest value (37,3%) in the case of Pichiaceae and the largest values (80% and 70%) in Sporopachydermia clade and Saccharomycetaceae of the KLE clade, respectively (supplementary table S5I.a, Supplementary Material online). Five phylogenetic clades, i.e., Debaryomycetaceae and Metschnikowiaceae (both comprising the CUG-Ser1 clade), Saccharomycetaceae (belonging to the ZT and post-WGD clades), Saccharomycodaceae, and Trichomonascaceae showed an IB intron distribution of approximately 50%. Introns of subtypes ID and IA were the next most commonly found ones within the yeast phylogenetic clades, in 16.16% and 11.74% of the genomes, respectively (supplementary table S5I.a, Supplementary Material online). Group II introns were located in only 10 species, which either contained more than one group II introns (i.e., *S. malanga* and *S. ludwigii* in which all group II introns were found in *cox1*) or as single occurrences in fungal species belonging to CUG-Ala clade and the “early diverging” lineages, like Lipomycetaceae, and Trigonopsidaceae families, and Sporopachydermia clade (supplementary table S5II, Supplementary Material online).

When examining the intron distribution in relation to the mt host genes and the family or phylogenetic clade in which these genes were located, it became obvious that most intron-rich genes were those of *cox1* of the post-WGD Saccharomycetaceae clade, followed by *cox1* of the Debaryomycetaceae family and *cox1* of the Dipodascaceae, with 116, 104, and 79 cases, respectively (supplementary table S5I.c, Supplementary Material online). Introns of Dipodascaceae and Trigonopsidaceae families were distributed in six genes, respectively, while in the rest of the families, introns were only located in one to five genes (supplementary table S5I.c, Supplementary Material online).



**Fig 4.**—Mean RSCU values (heatmaps) for each codon of (A) all families and (B) all phylogenetic groups. Mean RSCU values were calculated using the respective mean values of each genome, from all taxa belonging to each clade/phylogenetic group.

Intron insertion sites correlated with both their included ORF (homing endonuclease gene—HEG or reverse transcriptase gene) and their host gene, and they presented three different motifs of intronic distribution: 1) introns which, while carrying intronic ORF of similar type, may be found as ancestral elements at the same insertion site of a mt gene, 2) they may be mobile from one gene to another (due to similar insertion site), and 3) their intronic ORF may independently be mobile or lost (supplementary table S5, Supplementary Material online). Single species examples which covered all three mentioned observations could be detected in families Trigonopsidaceae, Lipomycetaceae, Dipodasceae, Trichomonasceae, and the CUG-Ser2 clade. In these mt genomes, the *cob* gene harbored an intron at position 429 bp. This intron belonged to subtype ID and carried a LAGLIDADG HEG. It was inserted at a conserved position (5' WYWNTGRGGT ↓ GCWACWGTWA 3') for all species examined (supplementary table S5II, Supplementary Material online). As an exception, in three *Yarrowia* species (*Y. deformans*, *Y. galli*, and *Y. lipolytica*), a similar intron ID carrying the LAGLIDADG gene was found at a similar insertion site of *cox3* (position 444 or 528 bp of gene). Moreover, in the case of *Y. galli* strain CBS9722, the third hypothesis related to motifs of intronic distribution could be confirmed, as at the same insertion position of the *cob* gene an ID intron was located, but it was ORF-less (supplementary table S5II, Supplementary Material online).

### GC clusters

The analysis of GC-clusters showed that 71 of the mitogenomes did not contain any GC repeats that met the applied search requirements (see Materials and Methods), while 49 of the genomes had less than 10 repeats and the remaining 35 had repeats ranging from 10 to 435 (*Y. prachuapensis* WB15 presented the maximum number of 435 repeats) (supplementary table S1, Supplementary Material online). The number of GC rich repeats seemed to be positively related to the size of the mitogenome ( $p < 0.001$ ,  $\rho = 0.5350$ ) according to the Spearman correlation test, while it was not related to GC content ( $p$ -value  $< 0.001$ ,  $\rho = -0.4328$ ). The examination of these repeats for each species individually showed sequence conservation and revealed the formation of distinct groups in each different organism based on this sequence similarity. Representative sequences, one of each GC clusters from all species, were aligned and showed that there is a highly conserved core of GC nucleotides in most of the species that had more than 10 GC rich repeats (supplementary file S3, Supplementary Material online).

### Discussion

Mt genomes analyses provide phylogenetic information, which, in a few deep-branched topologies, is different from the one revealed by the respective nuclear (nc)

analyses. The most characteristic difference is the alternate phylogenetic relation of *Saccharomycotina* with Taphrinomycotina and Pezizomycotina (Bullerwell et al. 2003; Pantou et al. 2008). However, in this study the phylogeny of the mt-based matrix (Fig. 1) was in accordance with the respective nc-based one (Shen et al. 2018). Both mt and nc datasets provided similar topologies with excellent statistical support. The nc-based phylogeny proposed that the CUG-Ser2 clade is positioned basally to Saccharomycetaceae and Phaffomycetaceae families and thus, the CUG-Ser2 clade is not phylogenetically related to the CUG-Ser1 clade. The results of our study further support that this phylogenetic distribution of CUG-Ser families was formed by two different evolutionary events (Shen et al. 2018) and support the hypothesis that the corresponding different reassignment of the genetic code for the CUG codon happened twice independently (Shen et al. 2018). Similarly, the phylogeny based on the mt data of this work is in agreement with the recent revision (see Schoch et al. 2020) placing CUG-Ala species (*Peterozyma toletana*, *P. tannophilus*, *Nakazawaea holstii*) as a monophyletic sister clade to the Pichiaceae family. Moreover, this work showed that the mt gene order of these species is conserved among them, and unique compared to the other syntenic groups. Thus, it is indicated that the mt sequences of CUG-Ala clade have a common mt ancestor, further supporting the argument that the CUG-Ala clade is a new phylogenetic and taxonomic group (Krassowski et al. 2018; Shen et al. 2018). Even though both phylogenies, produced by different matrices, are crucial for the unraveling of *Saccharomycotina* evolution, mitogenome based phylogenetic analysis is obtained faster and is equally reliable as the nuclear based one, since it resulted from the concatenation of only 14 mt proteins, whereas the nc dataset included 2,408 amino acid orthologous groups (Shen et al. 2018). Thus, it becomes evident that mitogenome based phylogenies are highly informative for providing accurate and easily generated phylogenies.

It is widely accepted that mt phylogenies present a different divergence rate when compared to the nc based ones, i.e., faster for metazoans, but slower for plants (Sandor et al. 2018). However, until now the different evolutionary rate in fungi has not been examined in depth, to our knowledge. Thus, in this work, the obtained rate is compared to the respective rate of the nuclear (nc) encoded dataset (supplementary file S1, Supplementary Material online). Mt based analyses presented a slower rate of divergence compared to the respective values of nuclear datasets during the early evolution of yeasts, but in later evolutionary stages, the opposite was observed. For instance, Pichiaceae and CUG-Ala clades diverged from CUG-Ser1 (and *H. uvarum*) at a time of 224/263 and 210 MYA based on the mt and nc datasets, respectively, while the split of the Saccharomycetaceae ZT clade from the post-WGD

species was set to 74/83 and 93 MYA from the mt and nc datasets, respectively (Fig. 2 and Shen et al. 2018).

In previous studies that tried to explain the codon reassignments in mt genetic codes of eukaryotes, several mechanisms have been suggested to explain these alterations (Sengupta and Higgs 2005, 2015). In detail, four mechanisms have been proposed, i.e., “Codon disappearance”, the “unassigned codon”, the “ambiguous intermediate” and the “compensatory change” (Sengupta et al. 2007), with the first two recently receiving more acceptance (Koonin and Novozhilov 2009; Sengupta and Higgs 2015). In this study, it became evident that the “Codon disappearance” mechanism might have driven evolution towards the observed genetic code changes through mutation and selection, as several codons like ATA, CTN, and TGG, are underrepresented, as shown by the RCSU index in this work (supplementary table S4, Supplementary Material online) and another study (LaBella et al. 2019). A comparison with the mt-based phylogenetic tree is needed to fully interpret genetic code alterations. Results from this study are in accordance with previous data from Sengupta et al. (2007) that described the major mt genetic code alteration events that took place during the evolution of all main phylogenetic eukaryotic lineages. The molecular clock obtained in our work provides an additional approach to estimate the timeframe during which the mt genetic code changed. In detail, the reassignment of a UGA stop to the Trp codon occurred as two independent events in the Ascomycota lineage (Sengupta et al. 2007), but only once during *Saccharomycotina* evolution (Fig. 2). This latter change in yeasts occurred 471/567 MYA, when the subphylum *Saccharomycotina* diverged from Taphrinomycotina (used as outgroup in this work). This codon reassignment also marked the deviation from the standard genetic code, since all yeasts examined besides the ones belonging to the Taphrinomycotina subphylum, employ the alternative mt genetic codes NCBI\_3 and NCBI\_4 (supplementary table S1, Supplementary Material online). CTN reassignment Leu to Thr occurred as a unique event right before the emergence of the Saccharomycetaceae and Saccharomycodaceae clades which took place approximately 129/134 MYA and marked the important evolutionary event of genetic code alteration from genetic code NCBI\_4 to NCBI\_3 (Fig. 2). This is further confirmed by comparing the high RSCU values of the respective CTN codons (genetic code NCBI\_3) among all phylogenetic groups (Fig. 4). CTN codon employment is more distinct in cases of CTT and CTA codons, since they present significantly higher RSCU values in the above-mentioned phylogenetic clades, compared with the majority of the remaining groups/families which employ the ACN codon for Threonine translation (Fig. 4). Reassignments of ATA from Ile to Met occurred as two independent events throughout yeast evolution: 1) the first event took place approximately 129/134 MYA within the KLE clade, right before the separation of

the subclade consisting of *Eremothecium* spp., the rest of the members of the KLE clade were not subjected to that genetic code reassignment (Fig. 2) and the second event took place within the post-WGD clade of the Saccharomycetes, after its separation from the ZT clade (Fig. 2). Sengupta et al. (2007) placed this event in the beginning of post-WGD clade formation in order to exclude members of the KLE clade, but did not analyze data from any representative of the ZT clade; thus, no conclusions could be made regarding codon reassignment within this clade, in that study. However, the reassignments of the genetic codes as presented in this study are supported because nowadays, information on mt genomes of more species is available. The codon frequency analysis showed a very low RSCU value of the ATA Met codon in the ZT clade, indicating that this genetic codon reassignment of Ile to Met did not occur in members of this clade. Thus, it may be assumed that this reassignment occurred at the beginning of the post-WGD clade formation in a relatively recent event, less than 74/83 MYA (Figs. 2 and 4).

Codon reassignments were not the sole important change during the evolution of mitogenomes, as was observed with the analysis of gene content and order. Mt genomes in all yeast species usually exhibit a conserved gene content but they also present a high size variability (Lang et al. 1999; Freel et al. 2015). This variability is analogous to that observed in genomes of the species belonging to Pezizomycotina, the other large subphylum of Ascomycota (Pantou et al. 2008; Megarioti et al. 2020). Our work showed that the range between the minimum and maximum known mt genome size in yeasts is now expanded to 11,223 bp (*H. pseudoguilliermondii*) and 173,850 bp (*S. malanga*). The size of the smallest genome can be attributed to a very low number of intergenic regions and the lack of introns and all *nad* genes, while the largest genome has a considerable number of introns and extremely large intergenic regions. However, total genome size fluctuations (supplementary table S1, Supplementary Material online) do not always coincide with the contribution of introns or intergenic regions, and there is either a tendency towards genome's size expansion, or towards size reduction as many different studies have shown (Wolters et al. 2015; Deng et al. 2018; De Chiara et al. 2020). Our analyses indicated the absence of a clear pattern during yeasts' mitogenome evolution (Fig. 2 and supplementary table S1, Supplementary Material online). Similarly, mt genome morphology (linear or circular mapping) is randomly found across the phylogenetic trees (Fig. 2). As a result, both size and morphology of yeast mt genomes may change in a ceaseless random manner throughout the species' evolution. The mt genome content, on the other hand, seems to be consistent in respect to its gene presence, throughout the evolution of yeasts, with the exception of the genes encoding NADH subunits and *rps3* (supplementary table S1, Supplementary Material online).

Gene order is intriguingly variable for yeasts. In this work, it became evident that the mt gene order of all species belonging to *Saccharomycotina* could split taxa into two highly and four less populated groups, respectively, presenting different distinct clusters (Fig. 3, Table 1 and supplementary table S3, Supplementary Material online). The correlation of the syntenic groups with the respective phylogeny of species involved, revealed that gene order not only follows, but may also affect phylogeny and to an extent also affects the evolution of the mt genomes. Some gene clusters were found intact in almost all syntenic groups but in a different sequential order, while other groups were unique to each phylogenetic clade (Fig. 3, Table 1). Nevertheless, the presence of specific small syntenic units, like *atp8-atp6*, is conserved in almost all of 155 strains examined. Moreover, conserved syntenic units of the early diverging yeasts, like *trnT-trnE-trnM1-trnM2*, can be also found in mt genomes of species belonging to other subphyla, such as Taphrinomycotina (this work) and Pezizomycotina (Kouvelis et al. 2004; Pantou et al. 2008). This designates the close relationship of these species with the Pezizomycotina subphylum. This observation agrees with previous phylogenetic studies which contributed to the examination of the synteny motifs in Pezizomycotina (Kouvelis et al. 2004; Pantou et al. 2008). If the above-mentioned previous studies are combined with the results presented in current work, it becomes evident that small syntenic units, like *nad4L-nad5*, *atp6-atp8*, and *trnS-rnl*, are consistent ancestral elements, present in the mt genome of the progenitor of Ascomycota. Furthermore, this syntenic conservation may be additionally explained through a functional approach, as was already shown in the case of Metazoa (Taanman 1999), where it was shown that small genes are co-translated in a polycistronic transcript. Therefore, the synteny analysis, as presented in Fig. 3 and supplementary fig. S3, Supplementary Material online, leads to the conclusion that a confined "relaxed" gene order is retained throughout yeast evolution.

The patterns of genome organization correlate with the tree topology. Members of the CUG-Ser 1 clade (Group B) exhibit a significant amount of genetic shuffling that took place approximately 213/241 MYA, during the alteration of genetic code. On the contrary, members of the Pichiaceae and CUG-Ala families (Groups C and D), which diverged from the CUG-Ser1 clade at a similar evolutionary time (242/263 MYA), present a smaller degree of gene shuffling and a partially different type of gene order and syntenic groups. In addition, members of group A, which was formed around the same period (approximately 227/241 MYA), showed a further differentiation in their gene order that coincides chronologically firstly with the discrimination of Phaffomycetaceae from Saccharomycetaceae, and secondly with the split of Saccharomycetaceae into three clades (post-WGD, ZT, and KLE).



Interestingly, the syntenic group *trnR-trnG-trnD-trnS* occurs in both Group F and Group A which diverged from their close relatives in 471 and 103 MYA, respectively, and cannot be detected at all in other species' genomes. This may be a result of two independent events, but may also indicate that this and other similar groups are remnants of a protoancestral origin.

Such an "organized" and "fragmented" synteny of mt genomes with conserved but mobile gene clusters is a later acquired characteristic during the evolution of the *Saccharomycotina* subphylum, at least in the majority of the phylogenetic clades. The diverse synteny of mitogenomes found for taxa belonging to the same subgroups and at inter-species level, indicates that this genome shuffling is an ongoing process.

The main mechanism for these gene shuffling events of several clusters across the *Saccharomycotina* subphylum indicates that the recombination events happen in a perpetual pace during evolution. Recombination requires repetitive identical sequences of variable size (depending on the type) and its results are usually the duplication or elimination of coding and non-coding regions. Scattered repetitions of enriched GC content in yeast mt genomes have been described as GC clusters (supplementary file S3, Supplementary Material online). They have been suggested as the substrate of genetic recombination in yeasts (de Zamaroczy and Bernardi 1986; Wu and Hao 2015). It was shown in this study that the GC repeats in several species presented a highly conserved core of GC nucleotides that is often reversed and complementary within their mt genome (supplementary file S3, Supplementary Material online). Thus, GC islands may be the cause of many recombination events (Dieckmann and Gandy 1987), as most of these GC clusters have been reported to be located in intergenic regions (de Zamaroczy and Bernardi 1986), but they cannot be the only cause of gene shuffling. The existence of *trn* genes at the bilateral ends of the described syntenic units (Fig. 3 and supplementary table S3, Supplementary Material online) may also contribute to genetic recombination and genome diversity. These *trn* genes at the boundaries of the syntenic units may form a potential hairpin-like secondary structure that may play the role of the repetitive elements needed for recombination, as already proposed for metazoa (Saccone et al. 1999) and Pezizomycotina (Pantou et al. 2008).

Finally, intron mobility has been proposed as another cause of genome diversity (Megarioti et al. 2020; Mukhopadhyay and Hausner 2021). In this work it was confirmed that there are some ancestral introns with conserved host gene and insertion site, such as the ID intron at position 429 bp of the *cob* gene (supplementary table S5, Supplementary Material online). This intron has also been detected in phylum Basidiomycota and in subphylum Pezizomycotina (Cinget and Bèlanger 2020; Theelen et al.

2021) and it has been linked to the regulation of *cob* maturation (Grasso et al. 2006; Vallières et al. 2011). According to our analyses of all introns included in the 155 mt genomes examined in this work, introns located in mt genomes of yeasts seem to be expanded, reduced, and transferred without any clear pattern. A result which is in accordance with the "aenaon" model suggested by Megarioti et al. (2020) where the restless perpetual coevolution of introns and HEGs was presented.

Several "back and forth" mt genome changes have been observed throughout the evolution of all yeast species as shown in this study (Fig. 2). While intron mobility and recombination are the two mechanisms which have been widely discussed as factors of gene shuffling in fungal mt genomes (Korovesi et al. 2018; Megarioti et al. 2020; Mukhopadhyay and Hausner 2021), there are numerous other mechanisms which could potentially contribute to this diversity found in yeast mt genomes, but these are not fully applicable based on the results of this work. In detail, it has been proposed that genetic drift is the main factor for size differentiation of mt genomes among yeasts belonging to the KLE clade (Xiao et al. 2017). Genetic drift is an incidental driven process without any clear direction, usually occurring in small-sized populations (Masel 2011). Therefore, while genetic drift cannot be precluded as a possible mechanism of mt genome diversity, genetic hitchhiking (or genetic draft) may be considered as another driving force as shown by Hill (2020). This hypothesis is supported in this work by the changes observed in partial gene clusters of mt genomes, as in several cases one gene entrains another (Fig. 3). However, genetic hitchhiking is most probably the result of genetic recombination and entails natural selection (Smith and Haigh 1974). As shown in this study, the absence of a clear pattern in syntenic changes of the mt genomes does not indicate that natural selection in yeasts may have played such a role in the observed mt genome diversity (Figs. 2 and 3). These arbitrary but perpetual changes found in yeast mt genomes seem similar to the proposed mechanism of selective sweeps that was proposed for mammals (Meiklejohn et al. 2007). This mechanism entails the influence of the external environment in shaping mt genomes through directional selection, as previously shown for Metazoa (Ballard et al. 2014; Barreto et al. 2018). Nevertheless, based on the results of this work, mt genetic recombination and intron mobility remain the key mechanisms for shaping mt genome diversity. Irrelevant to its type (i.e., homologous, non-homologous, site-specific recombination and transposition), genetic recombination may act as a solution to the cell's needs for genome repair to assemble a functional electron transport chain (Stein and Sia 2017). The restless random process of mt genome evolution resulting from intron mobility and recombination events can be explained if the usage of the "aenaon" model is not restricted to intron

mobility but broadened to gene shuffling within the mt genomes.

## Conclusions

In conclusion, the mt based phylogenetic analysis of *Saccharomycotina* yeasts representing 11 known families proved to be equally reliable, when compared to the respective nuclear genome-based analyses, with alignment of fewer genes needed. Moreover, the molecular time clock of the concatenated mt dataset employing two different methodologies showed in both cases that the evolution of yeasts is perpetual with different divergence rates, i.e., slower at the early stages of evolution and faster at later stages, when compared to nuclear based molecular dating. Mt genome divergence of yeasts follows the previously proposed “aenaeon” model which explains the intron mobility and its contribution to the mt genome diversity (Megarioti et al. 2020), since intron abundance and intergenic region variability of mitogenomes proved to be labile and major contributors to the genetic diversity of yeasts. This diversity is further attributed to the major rearrangements caused by genetic recombinational events but are also constrained by the conservation of the few ancestral syntenic units found throughout the main yeast phylogenetic lineages. Finally, genetic code alterations do happen throughout the evolution among the different families of yeasts, even during recent events, like the codon reassignment of ATA from Ile to Met, which occurred before the post-WGD clade formation. Therefore, mitogenomic analyses are crucial in deciphering the evolution of yeasts.

## Materials and methods

### Retrieval of DNA sequences

In this work, 175 mt genome sequences were examined. Of the 175 species, 155 belong to the subphylum of *Saccharomycotina* and 20 belong to the subphyla of Taphrinomycotina and Pezizomycotina which were used as outgroups. The 155 yeast species belong to 11 of the 12 *Saccharomycotina* families, since representative genomes for the 12th family were not available when the analyses were performed (last search November 01, 2021). The majority of the sequences (143 of the 175), including the species used as outgroups, were retrieved from complete mt sequence entries of GenBank. Eleven taxa were collected from various databanks, but a complete mt genome was not found and as a result, the precise gene limits for these taxa were not detected. For this reason, protein sequences were identified and retrieved using protein similarity searches with BLASTp (Altschul et al. 1990) and were further used in phylogenetic analyses (Table 2 and supplementary table S1, Supplementary Material online). Moreover, mt genomes of 23 taxa were extracted from

publicly available WGS projects as assembled contigs (supplementary table S6, Supplementary Material online), by using the tBLASTn/BLASTx/BLASTn tool for each WGS scaffold and comparing them with mt genome sequences of closely related species. In cases where mt genomes belonged to more than one contig different parts were assembled into one sequence using the Seqman tool of Lasergene Suite 11 (DNASTAR Inc., Madison, WI) (Burland 2000).

### Genome Annotation

To proceed with the mt genome analyses, annotation of all acquired genomes was verified manually. Protein coding and the ribosomal (rRNA) genes were identified using BLASTx and BLASTn (Altschul et al. 1990), respectively, after comparisons with related known sequences. The *trn* genes were detected using the online software tRNAScan-SE (Chan and Lowe 2019). The mt genomes of 32 species were re-annotated, since mis-annotations regarding gene limits and intron content were detected (referred as “modified” in supplementary table S6, Supplementary Material online). Additionally, 9 mt genomes were *de novo* annotated along with 23 genomes found in databanks as WGS projects (referred as “*de novo*” and “*de novo* from WGS” in supplementary table S6, Supplementary Material online, respectively).

### Genome analyses

Following the correction of genome annotations, comparative gene order (synteny) analyses were performed manually to locate any existing differentiations or similarities in genome organization among taxa belonging to different phylogenetic groups. To further extend our analyses, the total number of mt genes in each genome was recorded to detect any possible deletions or duplications, mostly of *trn* genes. Our study included genome size comparisons, as well as an analysis of intron abundance and their distribution among all genes. Intron characterization was performed using RNAweasel and MFannot (Lang et al. 2007). The *de novo* annotated mt genomes are available as supporting information in this manuscript as tabular format (supplementary table S7, Supplementary Material online) and as GenBank format (doi: 10.6084/m9.figshare.19714219). The transformation from the tabular format to GenBank format was performed in the web-based programme “Galaxy CPT Public” (Ramsey et al. 2020).

### Codon usage frequency

Alternative codon usage frequency was examined in all taxa included in this study. Fungal mt genomes utilize both alternative codes to the Universal Genetic Code (NCBI code 1), i.e., NCBI genetic code 3 (The Yeast mt genetic code) and 4 (The Mold, Protozoan, and Coelenterate Mitochondrial

**Table 2**

The families of *Saccharomycotina* examined in this study. The number of species/strains from each family is provided (details in [supplementary table S1, Supplementary Material](#) online)

Family	Number of Taxa	Phylogenetic Group (Clade)	Number of Taxa
Phaffomycetaceae	9	Phaffomycetaceae	9
Debaryomycetaceae	37	CUG-Ala	3
Metschnikowiaceae	4	CUG-Ser1	39
Saccharomycopsidaceae	1	CUG-Ser2	1
Trichomonascaceae (Dipodascaceae)	2	Dipodascaceae/Trichomosaceae	16
Dipodascaceae	12	Lipomycetaceae	7
Trichomonascaceae	2	Pichiaceae	11
Lipomycetaceae	7	Saccharomycetaceae KLE	16
Pichiaceae	12	Saccharomycetaceae post-WGD	27
Saccharomycetaceae	55	Saccharomycetaceae ZT	11
Saccharomycodaceae	10	Saccharomycodaceae	10
Trigonopsidaceae	4	Sporopachydermia	1
		Trigonopsidaceae	4
Total	155	Total	155

Code and the Mycoplasma/Spiroplasma Code), and thus, they present a few alternative codons encoding for the amino acids Threonine (T), Methionine (M), and Tryptophan (W). Codon usage frequency was examined using sequence manipulation suite (Stothard 2000), after inserting the gene sequence and the genetic code used by the species' mt genome. The related information regarding the mt genetic code utilized by each species was retrieved from the NCBI Taxonomy database. In case where codon frequency results were contradictory and unusually different from the rest, the mt genetic code was further examined and verified or even corrected in some species. Seven of the 14 protein coding genes were included in this study, since seven genes of NADH dehydrogenase subunits 1–6 and *nad4L* are not present in all studied fungal genomes, and thus would provide misleading evidence regarding comparative codon usage frequency of the whole mt genome. After listing the frequency of alternative codons, the RSCU index was calculated as an indication of possible codon usage bias (Sharp et al. 1986). The RSCU value for a codon reflects the observed frequency of that codon, divided by the expected frequency, assuming an equal usage of the synonymous codons for an amino acid (Sharp et al. 1986). The synonymous codons with RSCU values >1.0 have positive codon usage bias and are defined as abundant codons, whereas those with RSCU values <1.0 have negative codon usage bias and are defined as less-abundant codons.

In detail, the frequencies of synonymous codons within each amino acid class were calculated utilizing mt sequences of the protein coding genes *cob*, *atp6*, *atp8*, *atp9* and *cox1-3*. In the above analyses, codon frequencies and RSCU values were measured for each taxon and for each of the above genes. In addition, accumulative mean RSCU values were measured for 1) each codon and each gene separately, by adding the respective values of all

taxa (e.g., ATG of *cob* in all species) and dividing them with the total number of taxa ([supplementary table S4, Supplementary Material](#) online), 2) each codon in all genes of each taxon (e.g ATG of *cob* plus ATG of *cox1* of each strain etc.) divided with the number of genes ([supplementary table S4, Supplementary Material](#) online), and 3) a final mean RSCU value was calculated for each codon by summing RSCU values occurring in calculation and dividing them with the total number of taxa ([supplementary table S4, Supplementary Material](#) online). In all cases, RSCU values were measured separately for 1) species utilizing genetic code 3 and 4 and 2) for each codon within each family or phylogenetic cluster, respectively.

### GC cluster analyses

To investigate the role of concatenated GC nucleotides, all repeats present in the mt genomes of organisms belonging to the *Saccharomycotina* subphylum were initially detected using the UGene program (Okonechnikov et al. 2012), for each species separately. The minimum length of a GC nucleotide repeat was set to 30 bp and the maximum distance among repeats was increased to 20,000 bp compared to the 5,000 bp of the default parameters, in order to expand our search (more “relaxed” conditions). Repeat analyses were performed at 90% identity for the purposes of determining repeats. All other parameters were set to default. The repeats that contained less than 50% GC were discarded after which the repeats of each organism were clustered with CD-hit-test (Huang et al. 2010) with identity set to 80% and all other parameters at default. The representative sequence of each cluster was selected, and an alignment was performed with the representative sequences of all species using Clustal Omega (Madeira et al. 2019), and

the consensus of this alignment was used to create DNA logos with WebLogo (Crooks et al. 2004).

### Phylogenetic analyses

Amino acid sequences of all 14 protein coding genes usually found in mt genomes, i.e., Apocytochrome b (*cob*), Cytochrome c oxidases (*cox*) subunits 1-3, ATPase subunit 6,8,9 and NADH dehydrogenases (*nad*) subunits 1-6, *nad4L*, were collected in order to construct a concatenated matrix (supplementary file S4, Supplementary Material online). The collected amino acid sequences were aligned using ClustalW (Thompson et al. 1994) as implemented in Lasergene's MegAlign v.11 program (Burland 2000). Alignment parameters were set to default and the result was verified manually in order to ensure correct alignments. In species lacking the *nad* genes, the corresponding sites were replaced with the symbol of the missing information.

Phylogenetic trees were constructed using neighbor joining (NJ), Bayesian interference (BI) and maximum likelihood (ML) methods, using PAUP (Wilgenbusch and Swofford 2003), MrBayes (ver. 3.2) (Ronquist and Huelsenbeck 2003) and RAxML (Stamatakis 2014) plus IQ-TREE software (Nguyen et al. 2015), respectively. Ten Pezizomycotina species and 10 Taphrinomycotina species were used as outgroups. The methodology applied for NJ, BI, and ML phylogenies has already been described in previous studies (Korovesi et al. 2018; Kortsinoglou et al. 2019). In detail, for the NJ analyses, reliability of nodes was assessed using 1,000 bootstrap iterations for all the 14 individual and concatenated datasets. For the BI analyses the program ProtTest (ver. 3) (Abascal et al. 2005) was used in order to determine the model that was best suitable for our dataset. In the case of the concatenated dataset, the Bayesian Information Criterion was applied, and the best amino acid substitution model was found to be mtREV, with Gamma Distributed (G) rates among sites. Four independent MCMCMC searches were performed, using 10 M generations and sampling set adjusted every 100,000 generations. Convergence was checked visually by plotting likelihood scores vs. generation for the two runs. Burn-in was set to default in all cases. The ML analysis in IQ-TREE was implemented using the mixture substitution model mtREV with C60 in order to take under consideration the heterogeneity across sites and among sequences (Lartillot and Philippe 2004). Single gene phylogenetic analyses were also performed using the NJ method as previously mentioned (supplementary file S5, Supplementary Material online) and the sequence similarity and divergence of each gene between all examined taxa were calculated through Lasergene's MegAlign v.11 program (Burland 2000) (supplementary table S8, Supplementary Material online).

### Molecular dating

To estimate the divergence times among the subphylum *Saccharomycotina*, the concatenated tree, as produced from the Bayesian analysis, was used as input in the RelTime method of MEGA7 (Kumar et al. 2016) and the least square dating (LSD2) method of IQ-TREE (To et al. 2016). In both cases, four calibration points were used as follows: 1) *S. cerevisiae*—*S. uvarum* split (14.3-17.94 MYA), 2) *S. cerevisiae*—*Kluyveromyces lactis* split (103-106 MYA), 3) *S. cerevisiae*—*C. albicans* split (161-447 MYA), 4) *Saccharomycotina*—Taphrinomycotina (304-590 MYA) (Kumar et al. 2017; Shen et al. 2018). For the RelTime analysis, the mtREV amino acid substitution model with Gamma Distributed (G) rates among sites ( $\alpha = 0.6990$ ) was used. The estimated log likelihood value was  $-291342.08$ . For the LSD2 method, the same substitution model was used with default parameters. In order to determine the similarity of our estimated divergence dates in reference to estimations from nuclear genomes (Shen et al. 2018), the ensuing procedure was followed: at first was test whether the two methods used in this work provided comparable estimates. To achieve this, the divergence dates for each node provided by LSD2 method were subtracted from the dates of the corresponding nodes provided by RelTime method. A one-sample *t*-test, which was conducted despite non-normal distribution of data due to large sample size, showed that the dates estimated by the two programs were significantly different (Lumley et al. 2002). The RelTime method provides generally earlier divergence dates compared to LSD2 method. As such each estimation was compared separately to the nuclear estimates and the results were assessed to zero by *t*-test.

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

The authors wish to thank editor Prof. K.H. Wolfe and the two anonymous reviewers whose comments helped in improving the manuscript. This work was supported by "Special Account for Research Grants" of National and Kapodistrian University of Athens under Research Program (code no. 16673)

### Data Availability

The data underlying this article are available in the article and in its online supplementary material.

### Literature cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.



- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Ballard JWO, Pichaud N, Fox C. 2014. Mitochondrial DNA: more than an evolutionary bystander. *Funct Ecol.* 28:218–231.
- Ballard JW, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol.* 13(4):729–744.
- Barreto FS, et al. 2018. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat Ecol Evol.* 2(8):1250–1257.
- Bartelli TF, Ferreira RC, Colombo AL, Briones MR. 2013. Intraspecific comparative genomics of *Candida albicans* mitochondria reveals non-coding regions under neutral evolution. *Infect Genet Evol.* 14:302–312.
- Benítez T, Martínez P, Codón AC. 1996. Genetic constitution of industrial yeast. *Microbiologia* 12(3):371–384.
- Botstein D, Fink GR. 2011. Yeast: an experimental organism for 21st Century biology. *Genetics* 189(3):695–704.
- Bouchier C, Ma L, Créno S, Dujon B, Fairhead C. 2009. Complete mitochondrial genome sequences of three *Nakaseomyces* species reveal invasion by palindromic GC clusters and considerable size expansion. *FEMS Yeast Res.* 9(8):1283–1292.
- Brankovics B, et al. 2017. Mitochondrial genomes reveal recombination in the presumed asexual *Fusarium oxysporum* species complex. *BMC Genom.* 18(1):735.
- Braun EL. 2003. Innovation from reduction: gene loss, domain loss and sequence divergence in genome evolution. *Appl Bioinform.* 2(1):13–34.
- Bullerwell CE, Leigh J, Forget L, Lang BF. 2003. A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res.* 31(2):759–768.
- Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends Genet.* 19(12):709–716.
- Burland TG. 2000. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol.* 132:71–91.
- Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 1962:1–14.
- Cinget B, Bélanger RR. 2020. Discovery of new group I-D introns leads to creation of subtypes and link to an adaptive response of the mitochondrial genome in fungi. *RNA Biol.* 17(9):1252–1260.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14(6):1188–1190.
- De Chiara M, et al. 2020. Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biol.* 18(1):49.
- Deng Y, et al. 2018. Comparison of the mitochondrial genome sequences of six *Annulohyphoxylon stygium* isolates suggests short fragment insertions as a potential factor leading to larger genomic size. *Front Microbiol.* 9:2079.
- de Zamaroczy M, Bernardi G. 1985. Sequence organization of the mitochondrial genome of yeast—a review. *Gene* 37(1–3):1–17.
- de Zamaroczy M, Bernardi G. 1986. The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. *Gene* 41(1):1–22.
- Dieckmann CL, Gandy B. 1987. Preferential recombination between GC clusters in yeast mitochondrial DNA. *EMBO J.* 6(13):4197–4203.
- Dujon B. 2006. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet.* 22(7):375–387.
- Dujon B. 2010. Yeast evolutionary genomics. *Nat Rev Genet.* 11(7):512–524.
- Duò A, Bruggmann R, Zoller S, Bernt M, Grünig CR. 2012. Mitochondrial genome evolution in species belonging to the *Phialocephala fortinii* s.l. - *Acephala applanata* species complex. *BMC Genom.* 13:166.
- Eldarov MA, Kishkovskaia SA, Tanaschuk TN, Mardanov AV. 2016. Genomics and biochemistry of *Saccharomyces cerevisiae* wine yeast strains. *Biochemistry (Mosc)* 81(13):1650–1668.
- Escalera-Fanjul X, Quezada H, Riego-Ruiz L, González A. 2019. Whole-genome duplication and yeast's fruitful way of life. *Trends Genet.* 35(1):42–54.
- Fisher KJ, Buskirk SW, Vignogna RC, Marad DA, Lang GI, Otto SP. 2018. Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLoS Genet.* 14(5):e1007396.
- Foury F, Roganti T, Lecrenier N, Purnelle B. 1998. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* 440(3):325–331.
- Freel KC, Friedrich A, Schacherer J. 2015. Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.* 15(4):fov023.
- Fricova D, et al. 2010. The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5' termini. *Microbiology* 156(Pt 7):2153–2163.
- Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 198(2):755–771.
- Grasso V, Palermo S, Sierotzki H, Garibaldi A, Gisi U. 2006. Cytochrome b gene structure and consequences for resistance to Qo inhibitor fungicides in plant pathogens. *Pest Manag Sci.* 62(6):465–472.
- Hao W. 2022. From genome variation to molecular mechanisms: what we have learned from yeast mitochondrial genomes? *Front Microbiol.* 13:806575.
- Hatefi Y. 1985. The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem.* 54:1015–1069.
- Hill GE. 2020. Genetic hitchhiking, mitonuclear coadaptation, and the origins of mt DNA barcode gaps. *Ecol Evol.* 10(17):9048–9059.
- Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26(5):680–682.
- Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428(6983):617–624.
- Knop M. 2011. Yeast cell morphology and sexual reproduction—a short overview and some considerations. *C R Biol.* 334(8–9):599–606.
- Koonin EV, Novozhilov AS. 2009. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life.* 61(2):99–111.
- Korovesi AG, Ntertilis M, Kouvelis VN. 2018. Mt-*rps3* is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. *Mol Phylogenet Evol.* 127:74–86.
- Kortsinoglou AM, et al. 2019. The mitochondrial intergenic regions *nad1-cob* and *cob-rps3* as molecular identification tools for pathogenic members of the genus *Cryptococcus*. *FEMS Yeast Res.* 19(8):foz077.
- Kouvelis VN, Ghikas DV, Typas MA. 2004. The analysis of the complete mitochondrial genome of *Lecanicillium muscarium* (synonym *Verticillium lecanii*) suggests a minimum common gene organization in mtDNAs of Sordariomycetes: phylogenetic implications. *Fungal Genet Biol.* 41(10):930–940.
- Kouvelis VN, Sialakouma A, Typas MA. 2008. Mitochondrial gene sequences alone or combined with ITS region sequences provide firm molecular criteria for the classification of *Lecanicillium* species. *Mycol Res.* 112(Pt 7):829–844.
- Kováč L, Lazowska J, Slonimski PP. 1984. A yeast with linear molecules of mitochondrial DNA. *Mol Gen Genet.* 197(3):420–424.
- Krassowski T, et al. 2018. Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nat Commun.* 9(1):1887.
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 34(7):1812–1819.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.

- Kurtzman CP. 2011. Phylogeny of the ascomycetous yeasts and the re-naming of *Pichia anomala* to *Wickerhamomyces anomalus*. *Antonie Van Leeuwenhoek*. 99(1):13–23.
- Kurtzman CP, Boekhout T. 2017. Yeasts as distinct life forms of fungi. In: Buzzini P, Lachance MA and Yurkov A, editors. *Yeasts in natural ecosystems: ecology*. Cham: Springer.
- LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A, Barsh GS. 2019. Variation and selection on codon usage bias across an entire subphylum. *PLoS Genet*. 15(7):e1008304.
- Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet*. 33:351–397.
- Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet*. 23(3):119–125.
- Lartillot N, Philippe H. 2004 Jun. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21(6):1095–109.
- Lumley T, Diehr P, Emerson S, Chen L. 2002. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health*. 23:151–169.
- Madeira F, et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucl Acids Res*. 47(W1):W636–W641.
- Marcet-Houben M, Gabaldón T, Hurst LD. 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the Baker's yeast lineage. *PLoS Biol*. 13(8):e1002220.
- Masel J. 2011. Genetic drift. *Curr Biol*. 21(20):R837–R838.
- Megarioti AH, Kouvelis VN, Embley M. 2020. The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol Evol*. 12(8):1337–1354.
- Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. *Trends Genet*. 23(6):259–263.
- Miranda I, Silva R, Santos MA. 2006. Evolution of the genetic code in yeasts. *Yeast* 23(3):203–213.
- Mukhopadhyay J, Hausner G. 2021. Organellar introns in fungi, algae, and plants. *Cells* 10(8):2001.
- Nguyen NT, Schmidt HA, Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol*. 32:268–274.
- Nguyen DT, Wu B, Xiao S, Hao W. 2020. Evolution of a record-setting AT-rich genome: indel mutation, recombination, and substitution bias. *Genome Biol Evol*. 12(12):2344–2354.
- Nielsen J. 2019. Yeast systems biology: model organism and cell factory. *Biotechnol J*. 14(9):e1800421.
- Nosek J, Tomáška L. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr Genet*. 44(2):73–84.
- Nosek J, Tomáška L, Fukuhara H, Suyama Y, Kovác L. 1998. Linear mitochondrial genomes: 30 years down the line. *Trends Genet*. 14(5):184–188.
- Okonechnikov K, Golosova O, Fursov M. 2012. UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28(8):1166–1167.
- Pantou MP, Kouvelis VN, Typas MA. 2008. The complete mitochondrial genome of *Fusarium oxysporum*: insights into fungal mitochondrial evolution. *Gene* 419(1–2):7–15.
- Peter J, et al. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339–344.
- Pote ST, et al. 2020. Distribution of pathogenic yeasts in different clinical samples: their identification, antifungal susceptibility pattern, and cell invasion assays. *Infect Drug Resist*. 13:1133–1145.
- Pramateftaki PV, Kouvelis VN, Lanaridis P, Typas MA. 2006. The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. *FEMS Yeast Res*. 6(1):77–90.
- Ramsey J, et al. 2020. Galaxy and Apollo as a biologist-friendly interface for high-quality cooperative phage genome annotation. *PLoS Comput Biol*. 16(11):e1008214.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.
- Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* 238(1):195–209.
- Sandor S, Zhang Y, Xu J. 2018. Fungal mitochondrial genomes and genetic polymorphisms. *Appl Microbiol Biotechnol*. 102(22):9433–9448.
- Santos MA, Ueda T, Watanabe K, Tuite MF. 1997. The non-standard genetic code of *Candida* spp.: an evolving genetic code or a novel mechanism for adaptation? *Mol Microbiol*. 26(3):423–431.
- Schoch CL, et al. 2020. NCBI taxonomy: a comprehensive update on curation, resources and tools. Database: J Biol Databases Curation. 2020:baaa062.
- Sekito T, Okamoto K, Kitano H, Yoshida K. 1995. The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. *Curr Genet*. 28(1):39–53.
- Sengupta S, Higgs PG. 2005. A unified model of codon reassignment in alternative genetic codes. *Genetics* 170(2):831–840.
- Sengupta S, Higgs PG. 2015. Pathways of genetic code evolution in ancient and modern organisms. *J Mol Evol*. 80(5–6):229–243.
- Sengupta S, Yang X, Higgs PG. 2007. The mechanisms of codon reassignments in mitochondrial genetic codes. *J Mol Evol*. 64(6):662–688.
- Sharp PM, Tuohy TM, Mosurski KR. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucl Acids Res*. 14(13):5125–5143.
- Shen XX, et al. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 175(6):1533–1545.e20.
- Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res*. 23(1):23–35.
- Solieri L. 2010. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends Microbiol*. 18(11):521–530.
- Stajich JE, et al. 2009. The fungi. *Curr Biol*. 19(18):R840–R845.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stein A, Sia EA. 2017. Mitochondrial DNA repair and damage tolerance. *Front Biosci*. 22:920–943.
- Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28(6):1102–1104.
- Taanman JW. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta*. 1410(2):103–123.
- Teresa Fernández-Espinar M, Barrio E, Querol A. 2003. Analysis of the genetic variability in the species of the *Saccharomyces sensu stricto* complex. *Yeast* 20(14):1213–1226.
- Theelen B, Christinaki AC, Dawson TL, Boekhout T, Kouvelis VN. 2021. Comparative analysis of *Malassezia furfur* mitogenomes and the development of a mitochondria-based typing approach. *FEMS Yeast Res*. 21(7):foab051.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22(22):4673–4680.
- To TH, Jung M, Lycett S, Gascuel O. 2016 Jan. Fast dating using least-squares criteria and algorithms. *Syst Biol*. 65(1):82–97.
- Vallières C, Trouillard M, Dujardin G, Meunier B. 2011. Deleterious effect of the Qo inhibitor compound resistance-conferring mutation G143A in the intron-containing cytochrome b gene and mechanisms for bypassing it. *Appl Environ Microbiol*. 77(6):2088–2093.

- Wang JM, Bennett RJ, Anderson MZ, Nielsen K. 2018. The genome of the human pathogen *Candida albicans* is shaped by mutation and cryptic sexual recombination. *mBio* 9(5): e01205-18.
- Wesolowski M, Fukuhara H. 1981. Linear mitochondrial deoxyribonucleic acid from the yeast *Hansenula mrakii*. *Mol Cell Biol.* 1(5): 387–393.
- Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP\*. *Curr Protoc Bioinf.* doi:10.1002/0471250953.bi0604s00
- Williamson D. 2002. The curious history of yeast mitochondrial DNA. *Nat Rev Genet.* 3(6):475–481.
- Wolfe K. 2004. Evolutionary genomics: yeasts accelerate beyond BLAST. *Curr Biol.* 14(10):R392–R394.
- Wolfe KH. 2015. Origin of the yeast whole-genome duplication. *PLoS Biol.* 13(8):e1002221.
- Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634):708–713.
- Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genom.* 16(1):451.
- Wu B, Hao W. 2015. A dynamic mobile DNA family in the yeast mitochondrial genome. *G3 (Bethesda, Md.). G3 Genes Genomes Genet.* 5(6):1273–1282.
- Xiao S, Nguyen DT, Wu B, Hao W. 2017. Genetic drift and indel mutation in the evolution of yeast mitochondrial genome size. *Genome Biol Evol.* 9(11):3088–3099.

**Associate editor:** Kenneth Wolfe