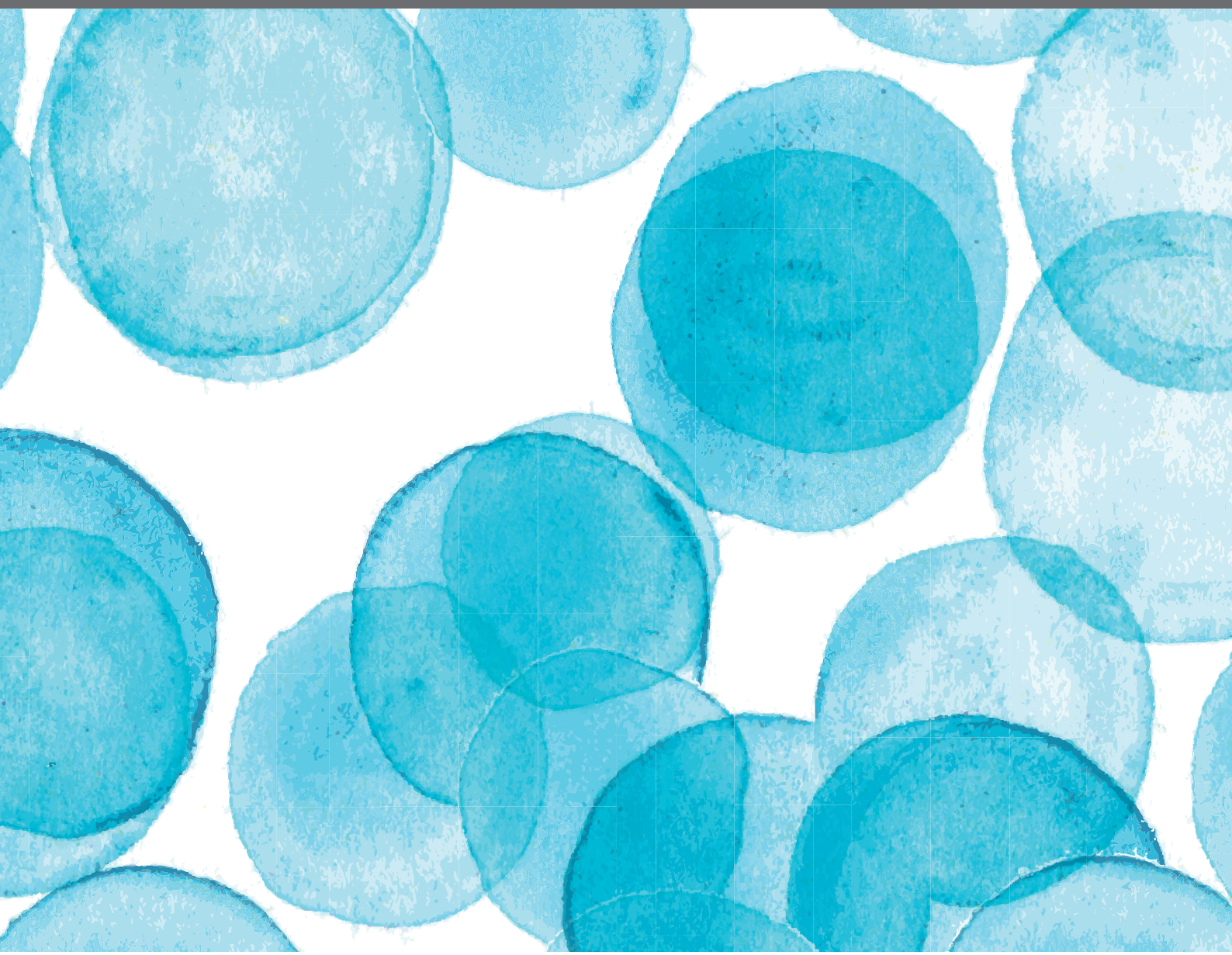
MITOCHONDRIAL GENOMES AND MITOCHONDRION RELATED GENE INSIGHTS TO FUNGAL EVOLUTION

EDITED BY: Vassili N. Kouvelis and Georg Hausner
PUBLISHED IN: Frontiers in Microbiology and Frontiers in Fungal Biology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-061-9
DOI 10.3389/978-2-88976-061-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MITOCHONDRIAL GENOMES AND MITOCHONDRION RELATED GENE INSIGHTS TO FUNGAL EVOLUTION

Topic Editors:
Vassili N. Kouvelis, National and Kapodistrian University of Athens, Greece
Georg Hausner, University of Manitoba, Canada

Citation: Kouvelis, V. N., Hausner, G., eds. (2022). Mitochondrial Genomes and Mitochondrion Related Gene Insights to Fungal Evolution.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-061-9

Table of Contents

frontiers | Frontiers in Microbiology

EDITORIAL
published: 11 April 2022
doi: 10.3389/fmicb.2022.897981

Editorial: Mitochondrial Genomes and Mitochondrion Related Gene Insights to Fungal Evolution

Vassili N. Kouvelis[1]* *and Georg Hausner*[2]

[1] Division of Genetics and Biotechnology, Department of Biology, School of Science, National and Kapodistrian University of Athens, Athens, Greece, [2] Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada

Keywords: fungi, organellar genomes, mitochondrial-nuclear interactions, mobile introns, mitogenomes

Editorial on the Research Topic:

Mitochondrial Genomes and Mitochondrion Related Gene Insights to Fungal Evolution

Mitochondria are organelles of eukaryotic cells that provide the platform for efficient energy metabolism, Fe/S-cluster biosynthesis, amino acid metabolism, and moreover, they have been linked to apoptosis, senescence, virulence, and drug resistance (Olson, 2001; Osiewacz et al., 2010; Chatre and Ricchetti, 2014; Giordano et al., 2018; Medina et al., 2020).

They originated from an ancestral α-proteobacterial endosymbiont (Margulis, 1970). Several complementary and alternative hypotheses to this firstly endosymbiotic theory have been proposed (for review see Martin et al., 2015 and references therein), which usually indicate with modifications the endosymbiosis of an α-proteobacterium within an archaeon. However, the additional participation of lysogenic viruses to the archaeal progenitor (Bell, 2009) or phage-like infected α-protebacterial progenitors (Varassas and Kouvelis) may have also contributed to the genesis of the proto-eukaryote. Mitochondria are semi-autonomous organelles, since they carry their own mitochondrial (mt) genomes and the components for protein synthesis. However, mitogenomes do not encode for all molecules necessary for the function and structure of this organelle. Maintenance of the mitogenome requires nuclear encoded factors that drive DNA replication, repair and transmission (Freel et al., 2015). Expression of mitochondrial genes is assumed to be regulated at the post-transcriptional level requiring nuclear encoded general and gene specific factors that guide transcription, RNA processing, intron splicing, RNA stability and translation (Lipinski et al., 2010; Varassas and Kouvelis). Mitogenome expression is linked with nuclear gene expression, establishing extensive inter-compartmental crosstalk that can integrate organellar gene expression into the cellular context as influenced by physiological, developmental, and environmental cues. A limited number of studies have shown mitonuclear interactions and more specifically, nuclear mitochondrial compatibility and co-adaptation, probably, are key components in fungal evolution and adaptation (Giordano et al., 2018; Steensels et al., 2021). Recently, Clergeot and Olson showed a link of nuclear and mitochondrial loci that affect radial growth of *Heterobasidion parviporum* heterokaryons (agent of root rot and butt rot of conifers); the mt involved gene has been identified as a unidentified ORF (uORF) (Himmelstrand et al., 2014), correlated to mt plasmids integrated to the mt genome (Medina et al., 2020). Mitogenomes probably encode uORFs and, by definition, these have no known function and homologs. These genes potentially evolved by endogenous events and although these might be viewed as accessory elements (or not essential), uORFs may have lineage specific functions that allow for fungi to adapt to certain environmental conditions or act as key drivers of evolution for host-pathogen interactions (Monteiro-Vitorello et al., 1995; Inoue et al., 2002; Patkar et al., 2012; van de Vossenberg et al., 2018).

Edited and reviewed by:
Ludmila Chistoserdova, University of Washington, United States

***Correspondence:**
Vassili N. Kouvelis
kouvelis@biol.uoa.gr

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *16 March 2022*
Accepted: *21 March 2022*
Published: *11 April 2022*

Citation:
Kouvelis VN and Hausner G (2022) Editorial: Mitochondrial Genomes and Mitochondrion Related Gene Insights to Fungal Evolution. Front. Microbiol. 13:897981. doi: 10.3389/fmicb.2022.897981

In general, fungal mitogenomes contain genes which encode products (RNAs or proteins) involved in translation (the small and large ribosomal subunit RNAs (*rns* and *rnl*) and a set of tRNAs), plus genes encoding protein components involved in the electron transport chain and oxidative phosphorylation. This includes parts of Complex I (subunits of NADH dehydrogenase: *nad*1 to *nad*6 and *nad*4L; except for members of the Taphrinomycota and Saccharomycetaceae and Saccharomycodaceae families of the Saccharomycetales), components of Complex III (*cob*) and Complex IV (*cox*1, *cox*2, and *cox*3), plus members of Complex V (ATP synthase components: *atp*6, *atp*8, and usually *atp*9) (Zardoya, 2020). Mitogenomes can also encode a ribosomal protein (*rps3* or *var1*) and the RNA (*rnp*B gene) component for RNaseP (Lang, 2018). The above-mentioned genes are designated as the conserved elements of fungal mitogenomes, even though it was recently shown that one or more of these genes may be also absent, arbitrarily in fungi independent to their taxonomic position (Korovesi et al., 2018; Fonseca et al.). Mitogenomes also include many accessory genes and elements in their content, besides the uORFs mentioned above. For example self-splicing introns, intron encoded ORFs, uORFs and in some members of the Ascomycota mitochondrial ORFs have been detected that appear to encode putative N-acetyltransferases and amino-transferases (Wai et al., 2019). Variability in mitogenome size is in part due to intergenic spacers, duplications, proliferation of repeats, and insertions of plasmid components or other elements (Bullerwell and Lang, 2005; Himmelstrand et al., 2014; Medina et al., 2020) (Fonseca et al.; Hao). All the above elements render fungal mitogenomes greatly diverse in content and ranging in size from 12.055 to > 500 kb (James et al., 2013; Liu et al., 2020).

Mt protein and rRNA coding genes are, usually, interrupted by introns that based on the RNA secondary structure and their splicing mechanisms can be assigned to either group I or group II introns (Michel and Westhof, 1990; Lang et al., 2007; Prince et al.). Mitochondrial introns are potentially self-splicing but to achieve splicing competent configurations they need to recruit protein factors (reviewed in Prince et al.). Organellar introns can be mobile elements as they encode intron-encoded proteins (IEPs) that may catalyze the movement of an intron from an intron-containing allele to cognate alleles that lack introns, a process referred to as intron homing or retro-homing, if mediated by reverse transcriptase activity (Belfort et al., 2002). Mobile introns (and their ORFs) are often referred to as diversity generating elements and they can be the major sources of mitogenome size polymorphisms within a species (Li et al.; Valenti et al.; Yildiz and Ozkilinc). However, there are examples where size variation and expansion are linked to repeats and not introns (Hao). In some fungal lineages, expansion of the mitochondrial genome is linked to the expansion of intron numbers (Megarioti and Kouvelis, 2020; Mukhopadhyay and Hausner, 2021), offering a possibility of fine tuning mitochondrial gene expression by nuclear factors that are involved in the splicing of group I and II introns (Rudan et al., 2018; Mukhopadhyay and Hausner, 2021; Lin et al.; Yildiz and Ozkilinic).

Mt accessory elements, like intergenic regions, where promoters, GC-clusters and other repetitive elements are located, show greater diversity and evolve faster, compared to the mt coding genes, which remain under purifying selection (Raffaele and Kamoun, 2012; Kolondra et al., 2015; Li et al.; Yildiz and Ozlilinc). Accessory elements can contribute to mt gene shuffling and the variable mitogenome reorganization through promoting recombinational events (Zhang et al.; Hao). This makes comparative mitogenome analyses essential in deciphering their evolution and diversity. In addition, this comparative analysis has been valuable in resolving issues related to fungal taxonomy, population genetics and diagnostics. On a global scale, fungal mitogenomes might be too variable to provide resolution to address some of the deeper phylogenetic issues within the Mycota (Fonseca et al.). As mentioned above, mitogenome architecture (gene composition and synteny) is highly variable among the fungi due to recombination events. These events are promoted by potential hyphal fusion associated with the existence of potential heteroplasmy (Zhang et al.). Combined with repeats promoting intrachromosomal recombination events and the potential horizontal movements of mobile elements (GC clusters, group I and II introns, homing endonuclease genes) plus uniparental inheritance, phylogenies based on mitogenomes have to be interpreted with caution when trying to address deeper phylogenetic questions (Aguileta et al., 2014; Stoddard, 2014; Repar and Warnecke, 2017; Mayers et al., 2021; Fonseca et al.; Hao). With regards to fungal pathogens, mitogenomic approaches have established potential links with fungicide/drug resistance and mitogenome features that can be linked to adaptation to specific hosts (Cinget and Bélanger, 2020; Wai and Hausner, 2021). On the latter issue, Lin et al. observed that among *Rhizoctonia solani* anastomosis groups there was some correlation between mitogenome gene expression patterns and the plant host, offering potential insights into fungal pathogens that have adapted to different hosts.

This special issue provides a cross section of research highlighting the various aspects of comparative mitogenomics and the potential of mitonuclear interactions on fungal adaptation and evolution. Yet, it also shows the need for more work on this topic, starting from improvements in accurate mitogenome annotations to the application of omics and systems biology approaches in unraveling the complexities of mitonuclear interactions and regulatory processes.

AUTHOR CONTRIBUTIONS

All authors drafted the Editorial and made direct and intellectual contributions to the work and approved the final version for publication.

FUNDING

This work was funded by Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery Grants Program (RGPIN-2020-05332 to GH).

REFERENCES

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., et al. (2014). High variability of mitochondrial gene order among fungi. *Genome Biol. Evol.* 6, 451–465. doi: 10.1093/gbe/evu028

Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B., and Lambowitz, A. M. (2002). "Mobile introns: pathways and proteins," in *Mobile DNA II*. eds N. L. Craig, R. Craigie, M. Gellert, and A. M. *Lambowitz (Washington, DC: ASM Press)*, 761–783.

Bell, P. J. (2009). The viral eukaryogenesis hypothesis: a key role for viruses in the emergence of eukaryotes from a prokaryotic world environment. *Ann N Y Acad Sci.* 1178, 91–105. doi: 10.1111/j.1749-6632.2009.04994.x

Bullerwell, C. E., and Lang, B. F. (2005). Fungal evolution: the case of the vanishing mitochondrion. *Curr. Opin. Microbiol.* 8, 362–369. doi: 10.1016/j.mib.2005.06.009

Chatre, L., and Ricchetti, M. (2014). Are mitochondria the Achilles' heel of the Kingdom Fungi? *Curr. Opin. Microbiol.* 20, 49–54. doi: 10.1016/j.mib.2014.05.001

Cinget, B., and Bélanger, R. R. (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, 1252–1260. doi: 10.1080/15476286.2020.1763024

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

Giordano, L., Sillo, F., Garbelotto, M., and Gonthier, P. (2018). Mitonuclear interactions may contribute to fitness of fungal hybrids. *Sci. Rep.* 8:1706. doi: 10.1038/s41598-018-19922-w

Himmelstrand, K., Olson, A., Brandström Durling, M., Karlsson, M., and Stenlid, J. (2014). Intronic and plasmid-derived regions contribute to the large mitochondrial genome sizes of Agaricomycetes. *Curr. Genet.* 60, 303–313. doi: 10.1007/s00294-014-0436-z

Inoue, I., Namiki, F., and Tsuge, T. (2002). Plant colonization by the vascular wilt fungus *Fusarium oxysporum* requires FOW1, a gene encoding a mitochondrial protein. *Plant Cell* 14, 1869–1883. doi: 10.1105/tpc.002576

James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., et al. (2013). Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr. Biol.* 23, 1548–1553. doi: 10.1016/j.cub.2013.06.057

Kolondra, A., Labedzka-Dmoch, K., Wenda, J. M., Drzewicka, K., and Golik, P. (2015). The transcriptome of *Candida albicans* mitochondria and the evolution of organellar transcription units in yeasts. *BMC Genom.* 16, 1–22. doi: 10.1186/s12864-015-2078-z

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. (2018). "Mitochondrial genomes in fungi," in *Molecular Life Sciences*, eds R. D. Wells, J. S. Bond, J. Klinman, and B. S. S. Masters (New York, NY: Springer-Verlag), 722–728. doi: 10.1007/978-1-4614-1531-2_113

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

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

Liu, W., Cai, Y., Zhang, Q., Shu, F., Chen, L., Ma, X., et al. (2020). Subchromosome-scale nuclear and complete mitochondrial genome characteristics of *Morchella crassipes*. *Int. J. Mol. Sci.* 21:483. doi: 10.3390/ijms21020483

Margulis, L. (1970). *Origin of eukaryotic cells*. New Haven, CT: Yale University Press.

Martin, W. F., Garg, S., and Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140330. doi: 10.1098/rstb.2014.0330

Mayers, C. G., Harrington, T. C., Wai, A., and Hausner, G. (2021). Recent and ongoing horizontal transfer of mitochondrial introns between two fungal tree pathogens. *Front. Microbiol.* 12:656609. doi: 10.3389/fmicb.2021.656609

Medina, R., Franco, M. E. E., Bartel, L. C., Martinez Alcántara, V., Saparrat, M. C. N., and Balatti, P. A. (2020). Fungal mitogenomes: relevant features to planning plant disease management. *Front. Microbiol.* 11:978. doi: 10.3389/fmicb.2020.00978

Megarioti, A. H., and Kouvelis, V. N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.* 12, 1337–1354. doi: 10.1093/gbe/evaa126

Michel, F., and Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216, 585–610. doi: 10.1016/0022-2836(90)90386-Z

Monteiro-Vitorello, C. B., Bell, J. A., Fulbright, D. W., and Bertrand, H. (1995). A cytoplasmically transmissible hypovirulence phenotype associated with mitochondrial DNA mutations in the chestnut blight fungus *Cryphonectria parasitica*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5935–5939. doi: 10.1073/pnas.92.13.5935

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

Olson, A, and Stenlid, J. (2001). Plant pathogens. *Mitochondrial control of fungal hybrid virulence. Nature* 411:438. doi: 10.1038/35078147

Osiewacz, H. D., Brust, D., Hamann, A., Kunstmann, B., Luce, K., Müller-Ohldach, M., et al. (2010). Mitochondrial pathways governing stress resistance, life, and death in the fungal aging model *Podospora anserina*. *Ann. N. Y. Acad. Sci.* 1197, 54–66. doi: 10.1111/j.1749-6632.2010.05190.x

Patkar, R. N., Ramos-Pamplona, M., Gupta, A. P., Fan, Y., and Naqvi, N. I. (2012). Mitochondrial b-oxidation regulates organellar integrity and is necessary for conidial germination and invasive growth in *Magnaporthe oryzae*. *Mol. Microbiol.* 86, 1345–1363. doi: 10.1111/mmi.12060

Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430. doi: 10.1038/nrmicro2790

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

Rudan, M., Dib, P. B., Musa, M., Kanunnikau, M., Sobočanec, S., Rueda, D., et al. (2018). Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *Elife* 7:e35330. doi: 10.7554/eLife.35330

Steensels, J., Gallone, B., and Verstrepen, K. J. (2021). Interspecific hybridization as a driver of fungal evolution and adaptation. *Nat. Rev. Microbiol.* 19, 485–500. doi: 10.1038/s41579-021-00537-4

Stoddard, B. L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* 5. doi: 10.1186/1759-8753-5-7

van de Vossenberg, B. T. L. H., Brankovics, B., Nguyen, H. D., 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

Wai, A., and Hausner, G. (2021). The mitochondrial genome of *Ophiostoma himal-ulmi* and comparison with other fungi causing Dutch elm disease. *Can. J. Microbiol.* 67, 584–598. doi: 10.1139/cjm-2020-0589

Wai, A., Shen, C., Carta, A., Dansen, A., Crous, P. W., and Hausner, G. (2019). Intron-encoded ribosomal proteins and N-acetyltransferases within the mitochondrial genomes of fungi: here today, gone tomorrow? *Mitochondrial DNA Part A DNA Mapping, Seq. Anal.* 30, 573–584. doi: 10.1080/24701394.2019.1580272

Zardoya, R. (2020). Recent advances in understanding mitochondrial genome diversity. *F1000Res.* 9:F1000 Faculty Rev-270. doi: 10.12688/f1000research.21490.1

Conflict of Interest: The 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 May 2021
doi: 10.3389/fmicb.2021.647989

Pan-Mitogenomics Approach Discovers Diversity and Dynamism in the Prominent Brown Rot Fungal Pathogens

*Gozde Yildiz[1] and Hilal Ozkilinc[1,2]**

[1] School of Graduate Studies, MSc Program in Biomolecular Sciences, Çanakkale Onsekiz Mart University, Çanakkale, Turkey, [2] Faculty of Arts and Sciences, Department of Molecular Biology and Genetics, Çanakkale Onsekiz Mart University, Çanakkale, Turkey

Monilinia fructicola and *Monilinia laxa* species are the most destructive and economically devastating fungal plant pathogens causing brown rot disease on stone and pome fruits worldwide. Mitochondrial genomes (mitogenomes) play critical roles influencing the mechanisms and directions of the evolution of fungal pathogens. The pan-mitogenomics approach predicts core and accessory regions of the mitochondrial genomes and explains the gain or loss of variation within and between species. The present study is a fungal pan-mitogenome of *M. fructicola* (N = 8) and *M. laxa* (N = 8) species. The completely sequenced and annotated mitogenomes showed high variability in size within and between the species. The mitogenomes of *M. laxa* were larger, ranging from 178,351 to 179,780bp, than the mitogenomes of *M. fructicola*, ranging from 158,607 to 167,838bp. However, size variation within the species showed that *M. fructicola* isolates were more variable in the size range than *M. laxa* isolates. All the mitogenomes included conserved mitochondrial genes, as well as variable regions including different mobile introns encoding homing endonucleases or maturase, non-coding introns, and repetitive elements. The linear model analysis supported the hypothesis that the mitogenome size expansion is due to presence of variable (accessory) regions. Gene synteny was mostly conserved among all samples, with the exception for order of the *rps3* in the mitogenome of one isolate. The mitogenomes presented AT richness; however, A/T and G/C skew varied among the mitochondrial genes. The purifying selection was detected in almost all the protein-coding genes (PCGs) between the species. However, *cytochrome b* was the only gene showing a positive selection signal among the total samples. Combined datasets of amino acid sequences of 14 core mitochondrial PCGs and *rps3* obtained from this study together with published mitochondrial genome sequences from some other species from Heliotales were used to infer a maximum likelihood (ML) phylogenetic tree. ML tree indicated that both *Monilinia* species highly diverged from each other as well as some other fungal species from the same order. Mitogenomes harbor much information about the evolution of fungal plant pathogens, which could be useful to predict pathogenic life strategies.

Keywords: pan-mitogenomics, mitogenome, evolution, *Monilinia* species, brown rot

OPEN ACCESS

Edited by:
Vassili N. Kouvelis, National and Kapodistrian University of Athens, Greece

Reviewed by:
Kelle Cardin Freel, University of Hawai'i at Mānoa, United States
Sotiris Amillis, National and Kapodistrian University of Athens, Greece

***Correspondence:**
Hilal Ozkilinc
hilalozkilinc@comu.edu.tr

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 December 2020*
Accepted: *29 March 2021*
Published: *12 May 2021*

Citation:
Yildiz G and Ozkilinc H (2021) Pan-Mitogenomics Approach Discovers Diversity and Dynamism in the Prominent Brown Rot Fungal Pathogens. Front. Microbiol. 12:647989. doi: 10.3389/fmicb.2021.647989

INTRODUCTION

Fungi are one of the most remarkable and diverse kingdom, with approximately 720,256 species compared with other eukaryotic organisms worldwide (Blackwell, 2011; Badotti et al., 2017). Fungal genomics exhibit important data for studies of adaptive behavior and evolutionary research due to their highly dynamic and fast-evolving features. High-throughput sequencing technologies have allowed for sequencing of the tremendous number of fungal nuclear genomes across many species, with a total of 2,599[1]. However, relatively limited data is available for whole mitochondrial genomes. For instance, only 793 mitochondrial genomes have been announced by NCBI Organelle Genome Resources[2]. Mitochondrial genomes present valuable information to explain both adaptative traits and the evolution of pathogens. Fungal mitochondrial genes can be targeted for plant disease management (Medina et al., 2020) and provide specific markers for population studies, as well as species diagnosis (Santamaria et al., 2009). Furthermore, mitogenome data contribute to expand information of fungal phylogenetics (Chen et al., 2019; Nie et al., 2019; Li Q. et al., 2020). Fungal mitochondrial genomes consist of highly conserved proteins and RNA encoding genes related to respiration and translation processes (Aguileta et al., 2014; Franco et al., 2017). Moreover, the presence of mitochondrial-encoded ribosomal protein genes, such as *rps3*, and its homologs (*var1* and *S5*) differs among fungal groups, and these genes may have been transferred to the nuclear genomes in different eukaryotic species (Bullerwell et al., 2000; Smits et al., 2007; Sethuraman et al., 2009; Korovesi et al., 2018; Yildiz and Ozkilinc, 2020). Furthermore, copy number, gene duplications, gain/loss of introns, and transposable and repetitive elements are the main factors causing mitogenome variations (Basse, 2010; Aguileta et al., 2014). Because of these factors, mitogenome sizes may vary within and among fungal taxonomic groups (Burger et al., 2003). Recent studies showed that homing endonucleases, such as GIY-YIG and LAGLIDADG families, play a significant role in shaping fungal genome structure and contribute to variations within and between species (Sandor et al., 2018; Kolesnikova et al., 2019; Yildiz and Ozkilinc, 2020).

Fungal mitogenomes have been evaluated for genomic features (Li Q. et al., 2020), comparative mitogenomics (McCarthy and Fitzpatrick, 2019), and pan-mitogenomics (Brankovics et al., 2018). The pan-genomic approaches, using a comparative genomics-based methodology to identify the core and accessory genomes or genomic regions, were applied on bacterial genomes at first (Tettelin et al., 2008). Core genomes tend to be conserved among strains, such as many housekeeping genes involved in translation, metabolism, and oligopeptide metabolism (McCarthy and Fitzpatrick, 2019). However, the accessory genome includes dispensable, variable, and "unessential regions," which may not be present in all strains or isolates within a clade (Torres et al., 2020). Thus, genome content can vary in distinct populations of a single fungal species, and the inventory of the variation at the genomic level in different isolates is crucial to characterize the complete set of accessory genes (Stajich, 2017). Two-speed genome evolution is referred to indicate compartmentalization of the fungal genomes as shared and slowly evolving regions as well as variable and fast-evolving regions (Dong et al., 2015; Bertazzoni et al., 2018; Torres et al., 2020). Pangenomics approach has recently resolved different fungal nuclear genomes (Kelly and Ward, 2018; McCarthy and Fitzpatrick, 2019; Badet et al., 2020); however, only few studies focused on mitochondrial genomes. For instance, sequencing of mitogenomes of *Aspergillus* and *Penicillium* species were analyzed by presenting core and accessory genes through comparative mitogenome analyses (Joardar et al., 2012). The mitogenome of phytopathogenic fungus *Fusarium graminearum* was analyzed considering the pan-mitochondrial genomics concept (Brankovics et al., 2018).

Monilinia species include phytopathogenic fungi that belong to the *Ascomycota* division. They cause brown rot disease on many stone and pome fruits, which results in severe economic losses around the world. *Monilinia laxa*, *Monilinia fructicola*, and *Monilinia fructigena* are the prevalent pathogenic species of the *Monilinia* genus causing this disease (Holb, 2006, 2008). The complete mitogenome of *M. laxa* was characterized by our previous study for the first time and presented higher content of mobile introns in comparison to some of the other phytopathogenic species from closely related genera (Yildiz and Ozkilinc, 2020). Thus, we expected high mitochondrial diversity within this pathogenic species and its relative, causing the same disease. This study aimed to uncover mitochondrial variations, by pan-mitogenomic approach, of the 16 mitogenomes from the two prominent and most abundant species (*M. fructicola* and *M. laxa*) that are known to cause brown rot disease. Mitogenomes were annotated as well as phylogenetics, and evolutionary selections were evaluated based on the protein-coding regions of the mitogenomes. This provides an essential foundation for future studies on population genetics, taxonomy, and crop protection strategies from the perspective of mitogenomics.

MATERIALS AND METHODS

Fungal Isolates and DNA Extraction

Isolates of *Monilinia* species were selected from the collection of Dr. Ozkilinc. Original isolates were long term stored at −20°C on Whatman filter papers no 1. Isolates were obtained from infected peach fruits from different orchards in six cities of Turkey (Ozkilinc et al., 2020). Sixteen isolates from a large collection of *Monilinia* samples were selected. The list of selected fungal pathogens of *M. fructicola* and *M. laxa* species is represented in **Table 1**. Selected isolates were grown from their original stored cultures on potato dextrose agar media at 23°C in the dark. Mycelia were transferred to potato dextrose broth and incubated at room temperature on a rotary shaker at 150 rpm for 5–7 days for genomic DNA isolation (Yildiz and Ozkilinc, 2020). Total DNA extractions were carried out using a commercial kit for fungi/yeast genomic DNA isolation (Norgen Cat. 27300, Canada), following the manufacturer's protocol. Concentration and purity of DNAs were assessed with a spectrophotometer

[1]https://www.ncbi.nlm.nih.gov/genome

[2]https://www.ncbi.nlm.nih.gov/genome/organelle/

TABLE 1 | The list of selected fungal isolates from *M. fructicola* and *M. laxa* species.

Isolate Code	Species	City/Country	Host species	Orchard number
B5-A4	*M. fructicola*	Çanakkale/Turkey	Peach	5
T-B1-A5	*M. fructicola*	Izmir/Turkey	Peach	10
Ti-B3-A2	*M. fructicola*	Izmir/Turkey	Peach	3
Ti-B3-A3-2	*M. fructicola*	Izmir/Turkey	Peach	3
Yolkenari-1	*M. fructicola*	Izmir/Turkey	Peach	Greengrocers
SC-B2-A4	*M. fructicola*	Samsun/Turkey	Peach	2
BG-B1-A4	*M. fructicola*	Bursa/Turkey	Peach	1
BG-B1-A17	*M. fructicola*	Bursa/Turkey	Peach	1
2B1-A5	*M. laxa*	Çanakkale/Turkey	Peach	1
T-B1-A4-2	*M. laxa*	Izmir/Turkey	Peach	1
Ni-B3-A2	*M. laxa*	Nigde/Turkey	Peach	3
MM-B2-A2	*M. laxa*	Mersin/Turkey	Peach	2
MM-B4-A4	*M. laxa*	Mersin/Turkey	Peach	4
MT-B1-A3-1	*M. laxa*	Mersin/Turkey	Peach	1
Yildirim-1	*M. laxa*	Bursa/Turkey	Peach	1
Yildirim-2-10th	*M. laxa*	Bursa/Turkey	Peach	–

(NanoQuant Infinite M200, Tecan) and a fluorometer (Qubit 3.0, Thermo Fisher Scientific, United States), then the DNAs were sent to an external service for Illumina-based library construction and short-read sequencing (Macrogen Inc., Next-Generation Sequencing Service, Geumcheon-gu, Seoul, South Korea).

Genome Sequences and *de novo* Assembly

The whole-genome sequence libraries of the 16 *Monilinia* spp. isolates were constructed using Illumina platform with TruSeq Nano kit to acquire paired-end 2 × 151 bp with 350-bp insert size, provided by Macrogen Inc., Next-Generation Sequencing Service. Adapters and low-quality reads were removed from raw data by using Trimmomatic 0.36 (Bolger et al., 2014) with the parameters as followed previously (Yildiz and Ozkilinc, 2020). The reads were evaluated to control the quality of sequences by using FastQC (Andrews, 2010), and the quality-checked data were used for further analysis. The mitogenomes were assembled *de novo* and extracted from the complete genome data using GetOrganelle v1.6.2 (Jin et al., 2019) with K-mer value: 105, including the SPAdes v3.6.2 (Bankevich et al., 2012) assembly program. QUAST reports presenting contig sizes, N50, and L50 were checked for the assembled contigs, including possible mitogenomes (Gurevich et al., 2013). The obtained mitogenomes for each isolate were visualized using the Geneious 9.1.8 program (Kearse et al., 2012). Mitogenomes represented by more than one contig were mapped by referring to the other completed mitogenomes using Geneious 9.1.8 (Kearse et al., 2012).

Mitochondrial Genome Annotations and Gene Orders

Core protein-coding genes (PCGs), ribosomal RNA, transfer RNA, and introns were annotated using the online server MFannot (Beck and Lang, 2010) and Mitos WebServer (Bernt et al., 2013). Mold/Protozoan/Coelenterate mitochondrial genetic code four was used for the annotation of the mitogenomes. Annotation of rRNA and tRNA genes was checked by using RNAweasel (Beck and Lang, 2010) and tRNAscan-SE 2.0 (Lowe and Chan, 2016), respectively. Hypothetical proteins, including ORFs and LAGLIDADG, GIY-YIG homing endonuclease families, were detected in intergenic regions using ORFinder in NCBI and Sequence Manipulation Suite: ORFFinder (Stothard, 2000), as well as the product of possible protein sequences, were checked by smart-blast in NCBI. Possible predicted mitochondrial genes were confirmed by a basic local alignment search tool using a nucleotide blast (BLASTN) in NCBI. The previously annotated mitogenome of *M. laxa* isolate (Ni-B3-A2) (accession number: MN881998) was used as a reference to check the annotations. Annotated gene arrangements were analyzed by using MAUVE 2.3.1 software (Darling et al., 2010). Moreover, duplicate regions in mitogenomes of isolate *M. fructicola* were investigated by using the Geneious 9.1.8 program (Kearse et al., 2012).

Identification of Repetitive Sequences

The repetitive elements in the mitogenomes were determined by using Tandem Repeats Finder (Benson, 1999). The repetitive sequences and their motifs were compared within and between the species.

Pan-Mitogenomics Analysis to Predict Conserved and Variable Regions Within Species

The percentages of the conserved and variable regions of the mitogenomes within each species were determined using bioinformatics tools Spine and AGEnt (Ozer et al., 2014). Additionally, R programming language with a deviance function from the stats package (R Core Team, 2013) was applied to interpret the statistical analysis of whether the intragenic intron sizes have contributed to differentiation in mitogenome sizes within the species. Thus, a linear correlation between the mitochondrial genome size (as the dependent variable) and intron length (as the independent variable) was tested based on the null hypothesis of positive correlation expectation.

Estimation of Codon Usage and Evolutionary Selection Patterns in Mitogenomes of *Monilinia* Species

The non-synonymous (Ka) and synonymous substitution rates (Ks) were calculated for all PCGs by using DnaSP v6.10.01 (Rozas et al., 2017). Since all the coding regions were almost the same within the species, the evolution rate was estimated on the total data set from both species. The strength of selection was inferred by considering that if the calculated ratio is equal to, greater than, or less than 1 indicates neutral evolution, positive (diversifying) selection, or purifying (negative) selection, respectively. Ka/Ks values for all protein-coding regions were visualized with ggplot in R programming language (R Core Team, 2013). The Relative Synonymous Codon Usage (RSCU) was obtained using MEGA 7 software (Kumar et al., 2016) and determined for all coding

regions. Furthermore, the nucleotide frequency of occurrence in each protein-coding gene (including the full length of the exons and introns) as well as in genes related to the ribosome (*rnl* and *rns*) was assessed for A/T and G/C asymmetry by using the following formulas:

$$\text{AT skew} = (A - T)/(A + T);\ \text{GC skew} = (G - C)/(G + C)$$

Phylogenetic Analysis Based on Amino Acid Sequences of Mitochondrial Protein-Coding Genes

Amino acid translation of the PCGs in the mitochondrial genomes of *M. fructicola* and *M. laxa* isolates were obtained based on the mitochondrial translation code data four using the Geneious 9.1.8 program (Kearse et al., 2012). The phylogenetic tree was constructed using a concatenated amino acid matrix of the 14 core mitochondrial genes and ribosomal protein of *M. fructicola* and *M. laxa* isolates. To strengthen the evolutional relationships between our data and other genera of the Helotiales, amino acid sequences of each of the PCGs and ribosomal protein were included from published mitochondrial genome data. Additional datasets were obtained from NCBI GenBank under the following accession numbers; KC832409.1 (*Botrytinia fuckeliana*), KJ434027.1 (*Sclerotinia borealis*), KT283062.1 (*Sclerotinia sclerotiorum*), KF169905.1 (*Glarea lozoyensis*), NC_015789.1 (*Phialocephala subalpine*), KF650572.1 (*Rhynchosporium agropyri*), KF650575.1 (*Rhynchosporium secalis*), KF650573.1 (*Rhynchosporium commune*), and KF650574.1 (*Rhynchosporium orthosporum*). The multiple protein sequences were concatenated by using the Geneious 9.1.8 program (Kearse et al., 2012) and aligned by ClustalW using MEGA software version 7 (Kumar et al., 2016). The maximum likelihood (ML) was constructed using RAxmlGUI v2.0 (Silvestro and Michalak, 2012) with 1,000 bootstrap replicates under BLOSUM62 substitutional matrix. The phylogenetic tree was visualized by FigTree v1. 4. program (Rambaut, 2012) and rooted at the midpoint.

RESULTS

Sequence Features of Mitogenomes of *Monilinia* Species

Most of the mitogenomes were extracted as one contig. However, the mitogenomes of five isolates (Yolkenari-1, Yildirim-2–10th, Ti-B3-A3-2, BG-B1-A17, and SC-B2-A4) were represented by four contigs. These contigs were mapped by using the other completed mitogenomes as reference. Statistics of QUAST reports for the assembled mitogenomes of 16 isolates of *Monilinia* species are provided in **Table 2**. After mapping, the mitogenome sizes ranged from 158,607 to 167,838 bp for *M. fructicola* and from 178,351 to 179,780 bp for *M. laxa.* Intraspecific length variations within isolates of *M. fructicola* were larger than the variations observed for the *M. laxa* isolates. An isolate of *M. laxa* (isolate code is MM-B2-A2) had the largest mitogenome size with 179,780 bp, while the isolate (coded as T-B1-A5) from *M. fructicola* species had the smallest mitogenome size with 158,607 bp. Total GC% content of all mitogenomes ranged between 30.0 and 31.1%.

Annotated Mitogenomes of Brown Rot Fungal Pathogens

The annotated mitogenome, with circular structure, of *M. fructicola* (BG-B1-A4) was chosen as representative sequence and submitted for the first time to the NCBI GenBank (accession number MT005827) (**Figure 1** and **Supplementary Table 1**).

TABLE 2 | QUAST report for the 16 mitogenomes of *Monilinia* spp.

Sample ID	Contigs	Largest contig	Total length	N50*	N75	L50**	L75	GC%	AT%	Species
MM-B2-A2	1	179780	179780	179780	179780	1	1	30.09	69.91	*M. laxa*
2B1-A5	1	178431	178431	178431	178431	1	1	30.04	69.96	*M. laxa*
MT-B1-A3-1	1	178421	178421	178421	178421	1	1	30.05	69.95	*M. laxa*
Yildirim-1	1	178358	178358	178358	178358	1	1	30.06	69.94	*M. laxa*
Ni-B3-A2	1	178357	178357	178357	178357	1	1	30.06	69.94	*M. laxa*
MM-B4-A4	1	178353	178353	178353	178353	1	1	30.05	69.95	*M. laxa*
T-B1-A4-2	1	178351	178351	178351	178351	1	1	30.05	69.95	*M. laxa*
Ti-B3-A2	1	167471	167471	167471	167471	1	1	31.11	68.89	*M. fructicola*
B5-A4	1	159656	159656	159656	159656	1	1	30.95	69.05	*M. fructicola*
BG-B1-A4	1	159648	159648	159648	159648	1	1	30.94	69.06	*M. fructicola*
T-B1-A5	1	158607	158607	158607	158607	1	1	30.95	69.05	*M. fructicola*
Yolkenari-1	4	69540	149868	59243	59243	2	2	31.19	68.81	*M. fructicola*
Ti-B3-A3-2	4	66346	146595	59242	59242	2	2	31.21	68.79	*M. fructicola*
Yildirim-2-10th	4	64887	142596	56701	56701	2	2	31.05	68.95	*M. laxa*
BG-B1-A17	4	64885	142593	56701	56701	2	2	31.05	68.95	*M. fructicola*
SC-B2-A4	4	64884	142592	56701	56701	2	2	31.05	68.95	*M. fructicola*

**N50: length is at least half (50%) of the total base content of the assembly.*
***L50: is the number of contigs equal to or longer than the N50 length.*

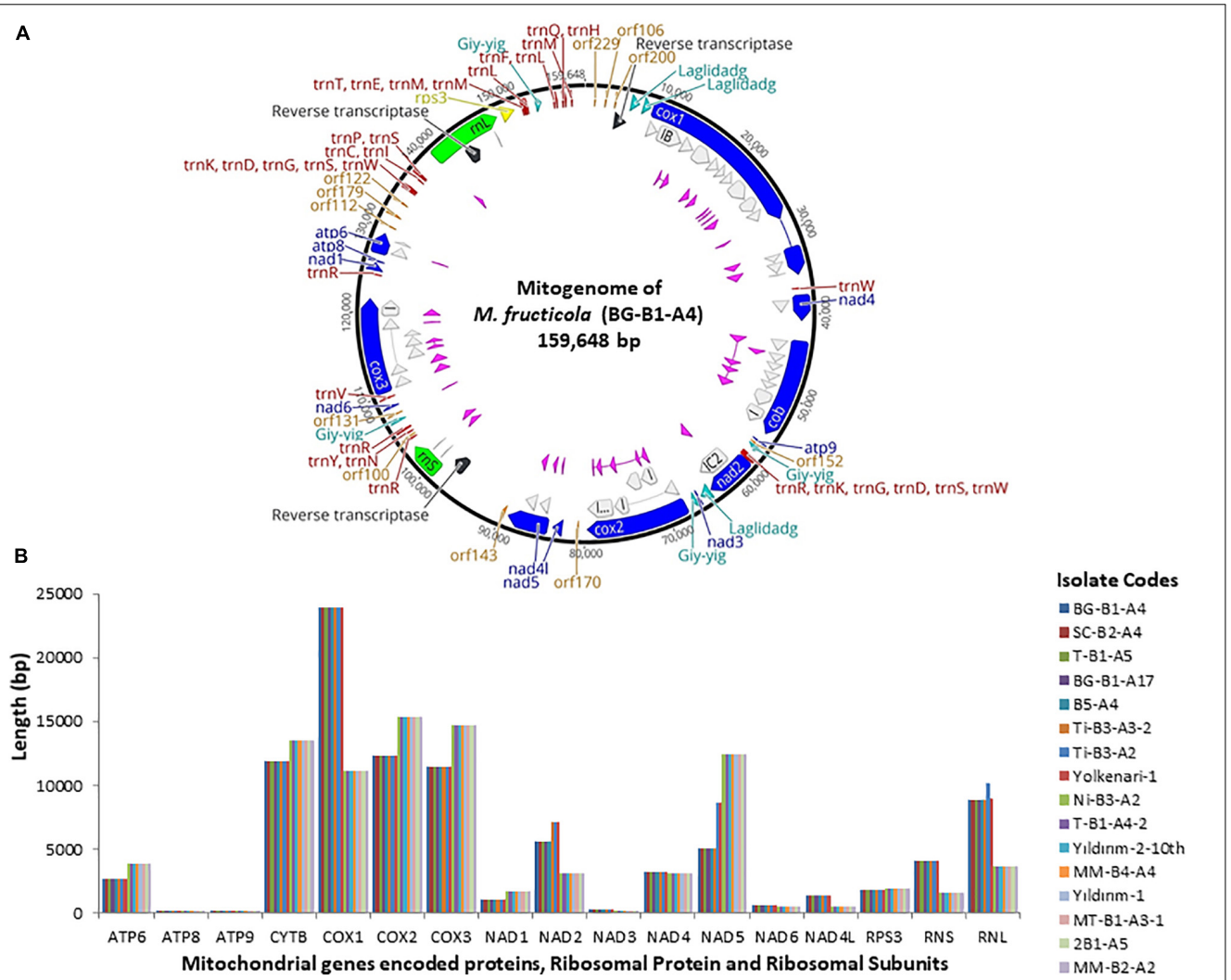


FIGURE 1 | (A) Annotated reference circular mitogenomes of *M. fructicola,* showing core conserved protein-coding genes (blue) with introns-encoded homing endonucleases (pink and turquoise), reverse transcriptase (black), orf (orange), two ribosomal subunits (green), rps3 (yellow), introns (gray), and tRNAs (red). **(B)** Length variation of 14 protein-coding genes, two ribosomal subunits, and rps3 among 16 mitogenomes of *Monilinia* pathogens.

Besides, all the mitogenomes were submitted to NCBI GenBank and the accession numbers were provided in the "Data Availability Statement" section at the end of the paper. On the other hand, the complete mitogenome of the isolate Ni-B3-A2 of *M. laxa* species was submitted to the NCBI GenBank with accession number MN881998 by our previous study (Yildiz and Ozkilinc, 2020). Mitogenomes of the 16 isolates of *Monilinia* species had the 14 PCGs responsible for mitochondrial oxidative phosphorylation system (OXPHOS) and ATP synthesis, including cytochrome c oxidase subunits 1, 2, and 3 (*cox1-3*), NADH dehydrogenase subunits 1–6 and 4L (*nad1-6* and *nad4l*), ATP synthase subunits 6, 8, and 9 (*atp6-8-9*), and cytochrome b (*cob* or *cytb)*. Furthermore, two ribosomal RNA genes for large and small subunits (*rnl* and *rns*), 32 transfer RNA (*tRNAs*), and a gene-encoding ribosomal protein S3 (*rps3*) were annotated. A set of 32 tRNAs of *Monilinia* isolates encoded for 20 essential amino acids were involved in the mitochondrial protein synthesis (**Figure 1**). However, alanine (Ala) and cysteine (Cys) amino acids were absent in the mitochondrial tRNAs of *M. fructicola* and *M. laxa*, respectively. Moreover, for both species, some of the tRNA genes encoding different anticodons corresponded to the same amino acids. For instance, four copies of arginine amino acid, which were encoded by TCT and TCG anticodons (**Supplementary Table 2**). The AT-rich content for tRNA codons was observed in both species (**Supplementary Table 2**). The mitogenomes of *M. fructicola* and *M. laxa* also contained some unidentified open reading frames, represented as ORFs encoding hypothetical proteins, and those ORFs were conserved within the species.

The largest gene sizes among the mitochondrial PCGs in the isolates of *M. fructicola* were detected in *cytochrome c oxidase* (*cox*) subunits *1,2,3* and *cytochrome b* (*cytb*), ranging from 23.8 to 11 kb (**Figure 1**). The large-sized PCGs of *M. laxa* isolates were *cox1,2,3*, *cytb*, and *nad5* genes, ranging from 15.3 to 11 kb. Gene

lengths were similar within the species but varied between the species. However, *atp8* and *atp9* genes showed the same size for all the 16, regardless of the species (**Figure 1** and **Supplementary Table 1**). An unknown sequence with 1,219-bp length was detected within the *rnl* gene, only in the mitogenomes of the isolates of *M. fructicola*. This unknown sequence was not matched with any sequence in the NCBI gene bank. Furthermore, another duplication of this unknown sequence was detected within the *rnl* of the isolate named Ti-B3-A2, which has one of the largest mitogenome among all. On the other hand, these sequences were not present in any of the isolates of *M. laxa*.

Skewness and Codon Usage Analysis

Nucleotide contents of the 16 mitogenomes were represented according to their AT and GC skew values (**Figure 2**). Many of the PCGs showed negative AT skews for both species, except for the genes of *atp6*, *cox1*, *cox2*, and *cox3*, which exhibited a positive AT skew (**Figure 2**). The AT skew of *nad4l* varied among species as positive and negative asymmetry for *M. fructicola* and *M. laxa*, respectively. The GC skews of the core PCGs for 16 mitogenomes showed positive asymmetry except for *atp8* and *nad3*, which exhibited negative GC skews. Moreover, the GC skew of *nad6* was positive for the isolates of *M. fructicola* but negative for the isolates of *M. laxa*. In addition, genes encoding ribosomal subunits (*rnl*, *rns)* and ribosomal protein (*rps3*) showed positive AT and GC skewness in all the mitogenomes of *Monilinia* samples (**Figure 2**).

Codon usage analysis for the 14 mitochondrial PCGs and *rps3* indicated that the most frequently used codons are as shown in the **Supplementary Table 3**. Codon usage patterns were quite similar between the 16 mitogenomes. The total number of codons was the highest for Leucine (Leu), Isoleucine (Ile), Lysine (Lys), and Phenylalanine (Phe) amino acids in mitogenomes of *Monilinia* species (**Supplementary Table 3**).

Comparison of Gene Arrangements of the Annotated Mitogenomes

The Mauve alignment reflected a conserved synteny among the 16 mitogenomes, which were divided into 11 homologous regions and represented with different colored synteny blocks (**Figure 3**). Gene order in the mitogenomes of *M. fructicola* isolates 3 was conserved with the exception for the order of the *rps3* gene, which

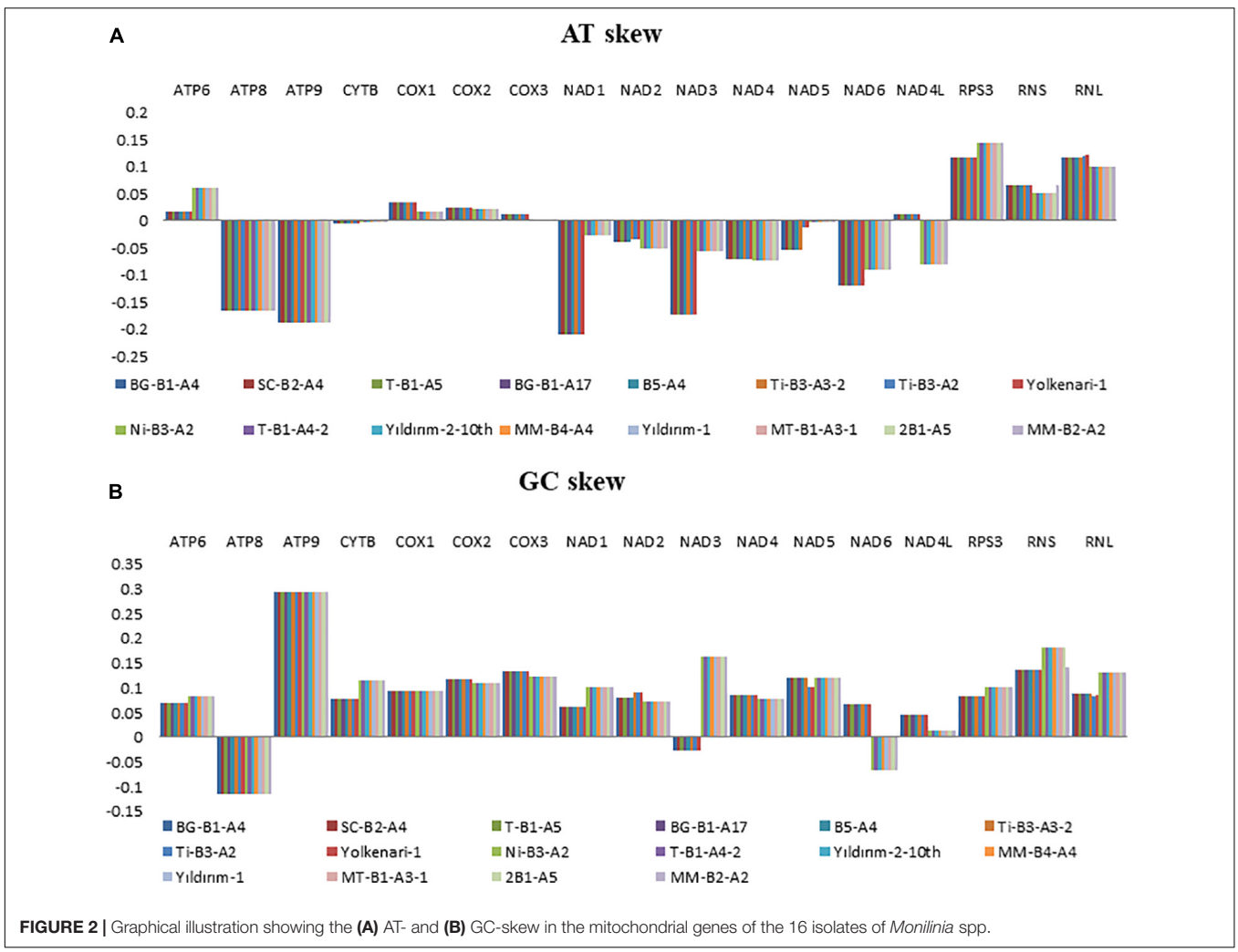


FIGURE 2 | Graphical illustration showing the **(A)** AT- and **(B)** GC-skew in the mitochondrial genes of the 16 isolates of *Monilinia* spp.

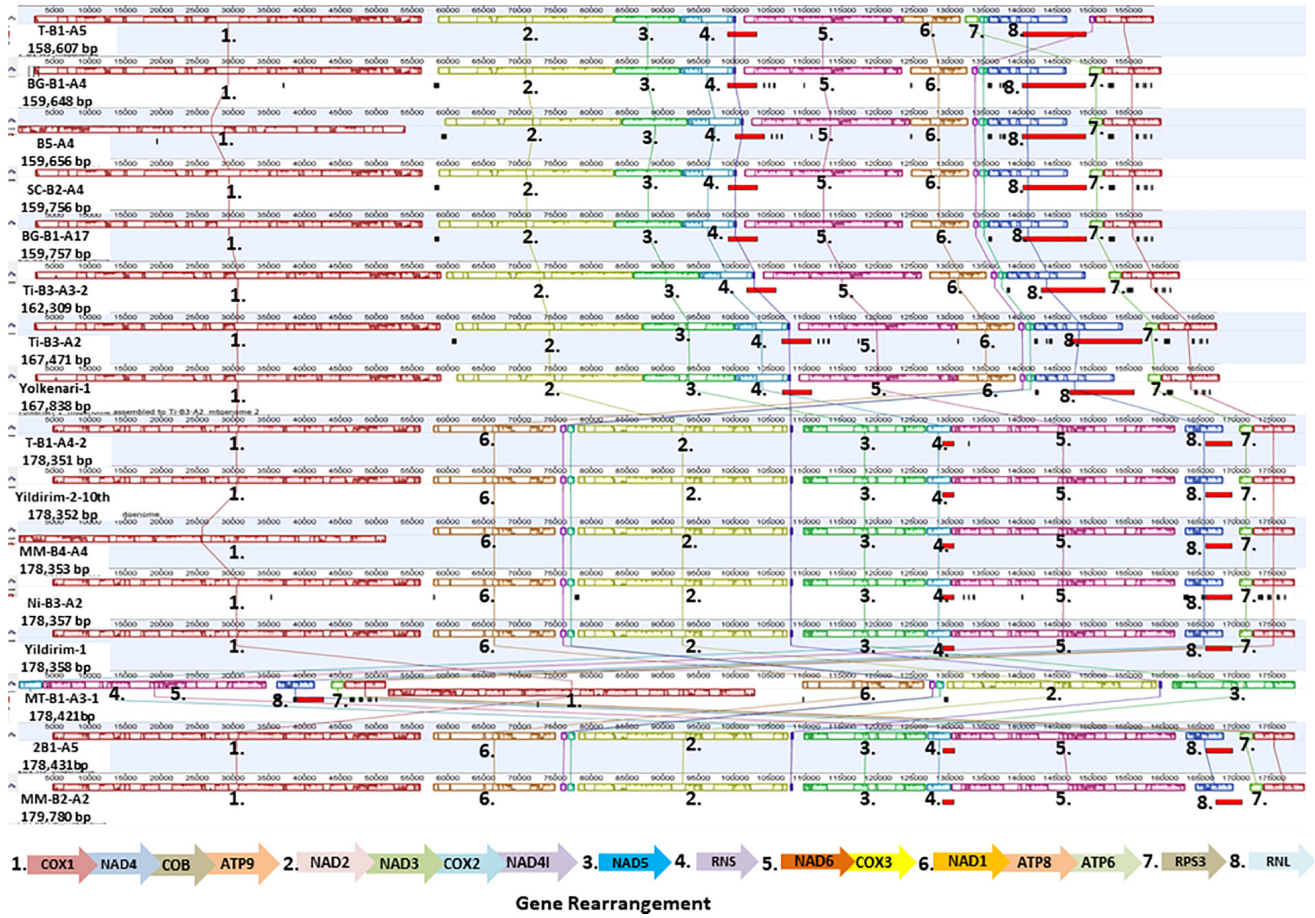


FIGURE 3 | Gene order of 16 mitogenomes of *Monilinia* isolates is shown. Homologous regions are colored differently and numbered to follow the order.

differed in the isolate coded T-B1-A5 (**Figure 3**). *Rps3* was found as a free-standing gene in the mitogenomes analyzed. Gene order in the mitogenomes of *M. laxa* isolates was conserved within the species (**Figure 3**).

Repetitive Sequences in the Mitogenomes of the Isolates of *Monilinia* species

Different repetitions were detected within the two species (**Table 3**). Isolates of *M. fructicola* presented two sequence motifs, which were $(TAC)_{29}$ and $(TC)_{18}$ located in the intergenic regions (**Table 3**). Among the eight mitogenomes of *M. fructicola*, 33–37 repeats were detected, and these repetitions covered 1.23–1.41% of the total mitogenome sizes. All the mitogenomes of *M. laxa* represented repetitive sequences in 59–60 bp in length comprising 1.60–1.69% of the total mitogenomes. The most longest repeats of more than 10 bp were $(AT)_{17}$ in the mitogenome of *M. laxa* was detected previously (Yildiz and Ozkilinc, 2020). This repetition was found in the seven other isolates of the species. $(AT)_{17}$ sequences were within an intron of the *cytb* gene, as reported previously (Yildiz and Ozkilinc, 2020).

Introns in the Mitogenomes of *Monilinia* Pathogens

Introns distributing within genic and intergenic regions, as well as mobile intron groups, were detected in all mitogenomes obtained from different isolates of two pathogenic species. The intron content of different genes varied in both species. In *M. fructicola*, the number of introns for each gene was *Cox1* with 13 introns, *cox2* with five introns, *cox3* with seven introns, *cytb* with seven introns, *nad2* and *nad4* with one intron, nad5 with two introns, atp6 with two introns, and large and small ribosomal subunit with four introns. On the other hand, *nad1*, *nad3*, *nad4l*, *atp8*, and *atp9* were found as intronless genes. Besides, some of the mitogenomes showed intron expansions within *nad2* and *nad5*. The isolates Ti-B3-A2, Ti-B3-A3-2, and Yolkenari-1 presented an additional intron (total 1,459 bp in size) in the *nad2* gene in comparison to the other samples. The *nad5* gene in the mitogenomes of the two isolates (Ti-B3-A2 and Yolkenari-1) included two extra introns (total 3,568 bp in size) to compare other isolates of the same species (**Figure 4**).

For all the eight mitogenomes of *M. laxa*, a total of 34 intron locations were found in *cox1* gene with six introns, *cox2* gene with seven introns, *cox3* gene with four introns, *cytb* gene with

TABLE 3 | Information on repetitive motifs detected in the mitogenomes of 16 isolates of number *Monilinia* spp.

Species	Isolate code	Repeat motif*	Copy number of the repeat**	Location of the repetition	Total repeats***
Monilinia laxa	2B1-A5	AT	17	Intron of cob gene	60
	MM-B2-A2	AT	17	Intron of cob gene	60
	T-B1-A4-2	AT	17	Intron of cob gene	60
	Yildirim-1	AT	17	Intron of cob gene	60
	Yildirim-2-10th	AT	17	Intron of cob gene	60
	Ni-B3-A2	AT	17	Intron of cob gene	60
	MM-B4-A4	AT	17	Intron of cob gene	59
	MT-B1-A3-1	AT	17	Intron of cob gene	59
Monilinia fructicola	B5-A4	TAC	29	Intergenic region	37
		TC	18		
	BG-B1-A17	TAC	29	Intergenic region	36
		TC	18		
	SC-B2-A4	TAC	29	Intergenic region	36
		TC	18		
	T-B1-A5	TAC	29	Intergenic region	36
		TC	18		
	Ti-B3-A2	TAC	29	Intergenic region	36
		TC	18		
	BG-B2-A4	TAC	29	Intergenic region	36
		TC	18		
	Yolkenari-1	TAC	29	Intergenic region	34
		TC	18		
	Ti-B3-A3-2	TAC	29	Intergenic region	33
		TC	18		

**The repeat motif having the highest copy number among the all repeats found in the mitogenome is shown.*
***Copy number of the whole repeat units is given.*
****All repeat units found in the mitogenome.*

seven introns, *nad5* gene with five introns, *nad1* gene with one intron, as well as the *nad2*, *nad4*, *atp6*, and *rnl* genes with one intron. On the other hand, *nad3*, *nad4l*, *atp8*, *atp9*, and *rns* did not contain any intron. Furthermore, MM-B2-A4 had one extra intron within the *rns* gene encoding a small ribosomal subunit. The total number of introns were greater in the mitogenomes of *M. fructicola* than in the mitogenomes of *M. laxa*; however, the total intron lengths were larger in the mitogenomes of *M. laxa* (**Figure 4**). Introns in the mitogenomes of *M. fructicola* isolates covered 42.2% of the whole mitogenome, and within the non-coding regions (28%) and within coding regions (14.2%). The introns in mitogenomes of *M. laxa* isolates covered 38.9% of the whole mitogenome and were within the non-coding region (16.3%), and within coding sequences (22.6%).

For *M. fructicola* isolates, the most intronic sequence carrying gene was the *cytb* with 72.8%, and intron rich genes followed by *cox1* (66.5%), *cox2* (61.5%), and *cox3* (53%). The most intronic content was found in the *cox2* gene with 65.4% and followed by *nad1* (65.2%), *cytb* (62.1%), and *cox3* (53%) for *M. laxa* isolates. Besides, group I and group II mobile introns were detected in the mitogenomes of *M. fructicola* (**Figure 1**). Different LAGDIDADG and GIY-YIG elements encoding homing endonucleases were detected within genic and intergenic regions of the mitogenomes of *M. fructicola*. Each of these elements from group I was represented as a single copy. Representation of group I mobile introns was given in detail for one isolate of *M. laxa* in our previous study (Yildiz and Ozkilinc, 2020). Group I mobile introns encoding homing endonucleases were found as approximately 18.1% of the whole mitogenome of *M. fructicola* and 35.4% of the whole mt-genome of *M. laxa*. Moreover, three different sequences were annotated as group II introns in the mitogenomes of *M. fructicola*. These sequences were the same and located in the same positions within all mitogenomes of the isolates of *M. fructicola*. These sequences were annotated as encoding reverse transcriptase/maturase (**Figure 1**). Group II introns were not detected within the mitogenomes of *M. laxa*.

Pan-Mitogenomics

The core mitochondrial genes included 14 PCGs, *rns*, *rnl*, and *rps3*. Except for *rnl*, *rns*, *nad2*, and *nad5*, all core genes were found fully conserved within species. Genic and intergenic introns, mobile introns (group I and group II), and repetitive sequences considered accessory regions of mitogenomes. However, many of these accessory elements were conserved within the species. The conserved regions formed a large portion of the mitogenome (ranging between 94 and 98%), while the variable regions covered 1.19–5.6% of the whole mitogenomes of the *M. fructicola* isolates. The most accessory-rich mitogenomes were detected in the isolates coded Ti-B3-A2 and Yolkenari-1 within *M. fructicola* (**Figure 5**). The whole mitogenomes of the isolates of *M. laxa* conserved within the species with the exception that one isolate (called MM-B4-A2) carried 0.7% of the total genome as a variable, which was not shared with any other isolate within this species (**Figure 5**).

According to the linear model test, intragenic introns had a significant effect ($P < 0.0001$ for both species) on mitogenome length variation. R^2 values explaining dependent (mitogenome length) and independent (intron length) variables were found as 0.9851 and 0.995 for *M. fructicola* and *M. laxa*, respectively.

Evolutionary Selection on Mitochondrial Genes of *Monilinia* Species

The evolutionary rates among the 16 mitogenomes from the two species showed that most of the genes were under negative (purifying selection) (**Figure 6**). The *Cox3* gene indicated neutral selection (**Figure 6**). The *cytb* gene was under diversifying or positive selection, and the remaining genes were under purifying selection (**Figure 6**).

Maximum Likelihood Analysis of Mitochondrial Protein-Coding Genes of *Monilinia* spp. and Other Genera From the Heliotales

The ML tree was obtained for the amino acid sequences of 14 PCGs and *rps3* of the *M. fructicola*, *M. laxa*, as well as some other species from Heliotales (**Figure 7**). Since all the amino acid sequences were conserved within the species, isolates were clustered together for each *Monilinia* species. However,

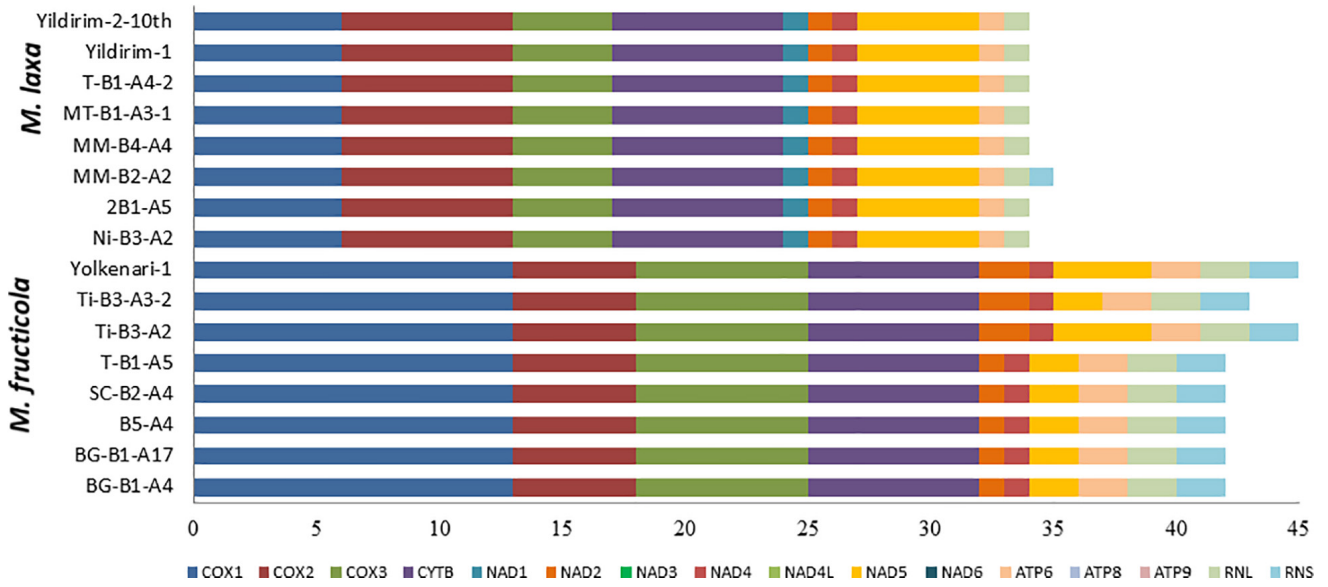


FIGURE 4 | Intron size variations among 16 isolates of *Monilinia* spp.

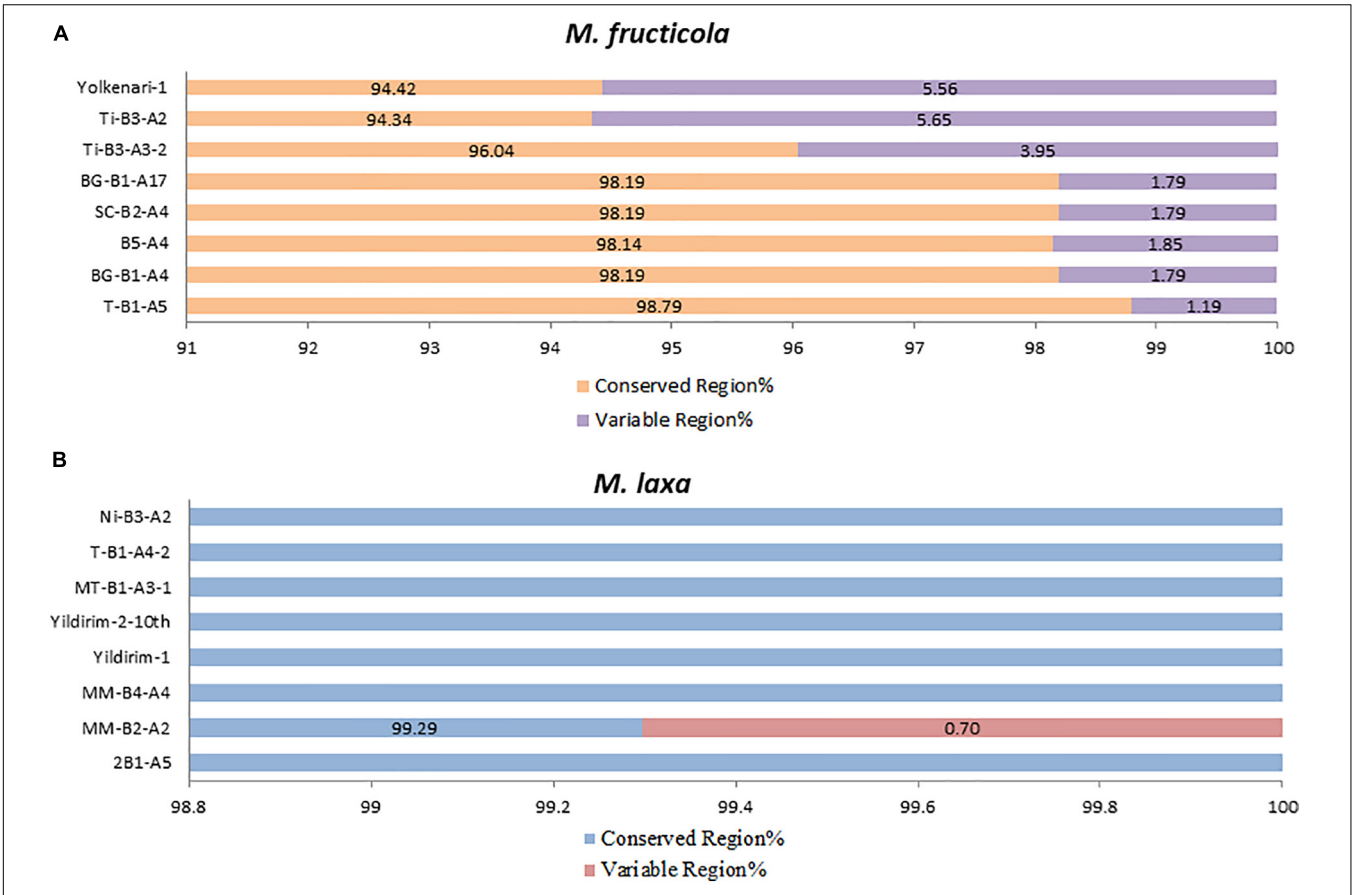


FIGURE 5 | Predicted percentage of conserved and variable regions of the mitogenomes within each species **(A)** *M. fructicola* and **(B)** *M. laxa*.

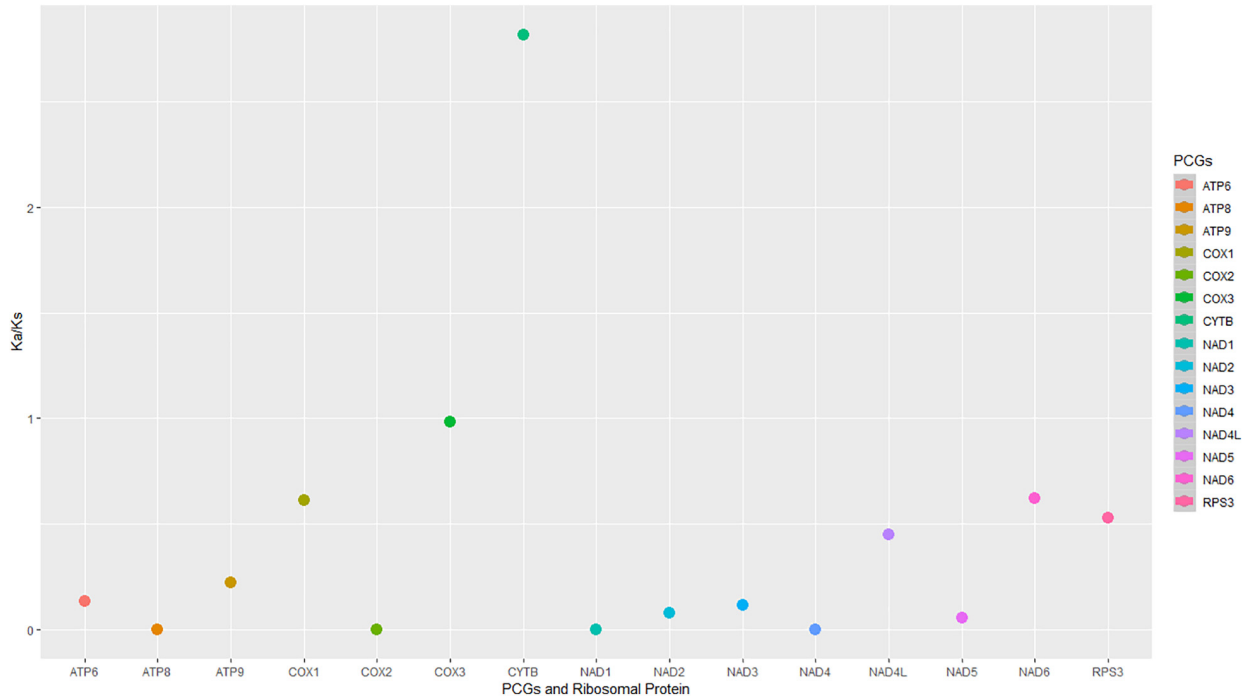


FIGURE 6 | The non-synonymous/synonymous mutation rates (Ka/Ks) were of the protein-coding regions and rps3 in the 16 mitogenomes from *M. fructicola* and *M. laxa* species.

M. fructicola and *M. laxa* diverged from each other with the high bootstrap value. Both *Monilinia* species were also distinctly related with the other species from Heliotales (**Figure 7**).

DISCUSSION

In this study, the complete mitogenomes of *M. fructicola* and *M. laxa* isolates were evaluated in-depth to understand variations within and between the species utilizing the pan-mitogenomic approach. The sizes of the 16 mitogenomes varied from 158,607 to 179,780 bp. The length of fungal mitogenomes is highly variable among the fungal species. It can range from 30 kb for the yeast *Candida parapsilosis* (Nosek et al., 2004) to 235 kb for *Rhizoctonia solani* (Losada et al., 2014). Based on the published data, mitogenomes of *Monilinia* pathogens seem to be quite large compared with other fungal mitogenomes. *S. borealis*, *S. sclerotiorum*, and *Botrytis cinerea*, closely related genera to *Monilinia*, presented mitogenomes of be 203, 128, and 82 kb in size, respectively (Mardanov et al., 2014). Mitogenome size differences are mainly related to the number and size of introns and repetitive elements that constitute the accessory part of the mitogenomes. Moreover, a variable number of genes of tRNAs, as well as loss/gain of some genes, such as *nad* and *atp* subunits, affect the mitogenome sizes (Aguileta et al., 2014; Franco et al., 2017; Li X. et al., 2020; Yildiz and Ozkilinc, 2020). In our study, the main factors that determined the differences in size of mitogenomes were introns, mobile intron groups, and repetitive sequences. Moreover, an unknown sequence and its duplicate was detected within the *rnl* gene within the isolates of *M. fructicola* and were not found in the mitogenomes of isolates of *M. laxa*. The origin and/or the duplication of this unknown sequence could be due to the mobility of mobile introns. Since this sequence was found within an rRNA-encoding gene, a group I intron-encoding RNA working as a ribozyme could be the source for this sequence. This further indicates that dynamism has been shaping the evolution and structure of the fungal mitogenomes continuously.

Fungal mitogenomes are known to have a high AT content, as confirmed in this study as well as in previous studies (Torriani et al., 2014; Franco et al., 2017). The high GC content of genomes was reported to affect the genome to evolve the advantage to maintain DNA stability in the high temperature, UV exposure, and fungicides (Musto et al., 2006; Marsolier-Kergoat, 2013). Furthermore, GC content has an important effect on evolutionary selection, recombination, gene conversion, and recombination in fungal plant pathogens (Stukenbrock and Dutheil, 2018). These considerations can be extended to mitogenomes as well.

Fungal mitogenomes have clustered with many tRNA genes with the different anticodons, indicating a strong preference for A or T, in the third position of codons. This strong preference using A/T has been defined in other species as wobble pairing and codon usage bias (Novoa and de Pouplana, 2012; Wei et al., 2014). However, having the decoders or iso-acceptors may cause mischarging (Pan, 2018), but this situation was not discussed or shown in any fungal mitochondrial genome.

Repetitive elements were 1.23–1.69% in the mitogenomes assessed in this study. The total number of these elements was

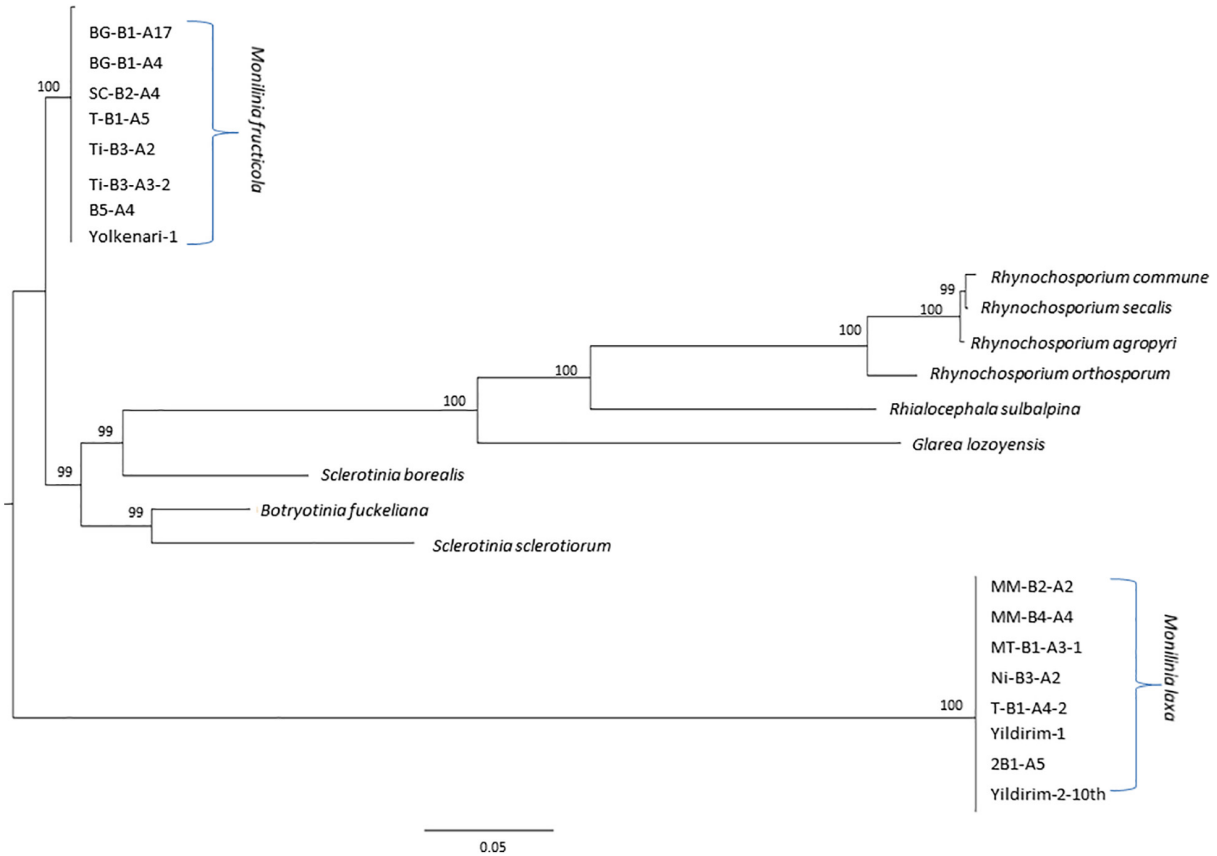


FIGURE 7 | Maximum likelihood tree inferred from the dataset of amino acid sequences of the 14 core mitochondrial genes and ribosomal protein for *Monilinia* isolates used in this study and other species of the Helotiales obtained from the NCBI GenBank. Node numbers show bootstrap support values. Concatenated order follows as *cox1*, *nad4*, *cob (cytb)*, *atp9*, *nad2*, *nad3*, *cox2*, *nad4l*, *nad5*, *nad6*, *cox3*, *nad1*, *atp8*, *atp6*, and *rps3*. Node numbers show bootstrap support values.

found greater and more diverse in *M. laxa* than in *M. fructicola*. Expansion of repetitive elements may have caused replication slippage and the correction of mitochondrial replication process together with proofreading efficiency may differ among the species. It is known that repeat-rich areas evolve more rapidly than other genomic regions (Raffaele and Kamoun, 2012; Dong et al., 2015). If the repeat-rich regions locate within genic regions, changes in these elements may indicate evolutionary selections of certain traits such as resistance or host adaptation (Raffaele et al., 2010; Nardi et al., 2012). Furthermore, exploring repetitive elements in mitogenomes could be highly useful in population genetics analyses, and they could have an important role for the dynamic structure of mitogenomes. Besides, mitochondrial repetitive elements can be useful molecular markers to study population structures.

Core PCGs related to mitochondrial OXPHOS and ATP synthesis are usually essential for the organisms' life and, thus, highly conserved within the mitogenomes. However, accessory regions could affect some traits, such as pathogenesis and virulence reference (Tettelin et al., 2008), and not be crucial for the survival of the organism. We observed a negative correlation between mitogenome size and virulence degree for these isolates (unpublished data by HÖ). It is well known that accessory regions in the nuclear genomes may change to adapt to evolutionary processes among the fungal isolates of the same species (Plissonneau et al., 2018; Badet and Croll, 2020). However, the effects of accessory regions of mitogenomes on different fungal traits have not been clarified. Mobile introns are another primary source of the size difference for mitogenomes that are highly found in both *Monilinia* species. Our previous study showed that *M. laxa* was the most intron-rich species compared with closely related species from other genera (Yildiz and Ozkilinc, 2020). These mobile introns included many different elements from group I in both species as well as group II introns in *M. fructicola*. Relatedness and phylogenetic signals of these elements within and between *Monilinia* species is another question that requires further investigations. Besides, these elements may have a contribution to certain traits such as virulence or fungicide resistance, and our ongoing studies have been directed to answer those questions.

Pan-mitogenomics approach identified that the core and accessory compartmentalizations occurred within *M. fructicola* species. Even though isolates of *M. laxa* carried many different introns, mobile elements, and repetitive sequences considered as an accessory part of the genomes, these sequences were mainly conserved among the mitogenomes within this species.

Only one isolate of *M. laxa* diverged from the other *M. laxa* isolates with a unique region that was represented as a variable part for the mitogenomes of *M. laxa*. This approach indicates that introns, mobile groups, and repetition patterns are highly conserved and stable within the mitogenomes of *M. laxa*. In contrast, the mitogenomes of *M. fructicola* showed variability and dynamism within the species. This could be related to possible recombination and/or selection pressures on mitogenomes of *M. fructicola*.

A/T and G/C skewness varied among the PCGs as well as between the species for some genes. Interestingly, genes related to ribosomal RNA and protein synthesis were in positive G/C and A/T skewness, indicating the richness of G and A over C and T, respectively. GC compositions were also interpreted as related to transcription start sites in plants and fungi (Fujimori et al., 2005). The different skewness along the regions may also indicate diverse selection pressures on two species. This study indicated that most of the coding genes have been evolving under strong purifying selection between the species according to the Ka/Ks ratios. Diversely, *cytb* was under a positive selection signal between the species. *Cytb* gene is one of the target regions of respiratory inhibitors that have been intensively used against fungal plant pathogens. Positive selection signals on this gene could be related to different adaptation responses of the species against fungicide selection pressure, which will be discussed in our further study for these pathogens.

Amino acid sequences of the core mitochondrial PCGs and *rps3* were fully conserved within *M. fructicola* and *M. laxa* species. Combined protein-coding data set based on ML phylogenetic indicated that these two species are highly diverged from each other as well as some from other fungal species from the Heliotales group. However, since one of the main variation contributors is the mobile introns, phylogenetic effects of these elements would be interesting. Mobile introns may shape evolutionary relationships differently in comparison with core PCGs as presented by Megarioti and Kouvelis (2020). Besides, mobile introns may reveal evolving lineages within each fungal species, and this will be investigated in a further study.

Considerable mitogenomic variations were observed within and between these two important pathogenic species within the *Monilinia* genus. Fungal mitochondrial genomes are still waiting for many hidden information on fungal traits and evolution. Pangenomic approach is successfully applicable for fungal mitogenomes due to relatively expanded accessory regions, as shown in this study. Since these organisms are important plant pathogens worldwide, resolving their mitogenomes may suggest new disease management strategies and predictions in evolutionary trajectories of the pathogens and the disease.

DATA AVAILABILITY STATEMENT

The first mitogenome of *M. fructicola* was submitted in this study can be found in GenBank (Accession number MT005827). Besides, the mitogenome of *M. laxa* used in this study, was submitted in our previous research and can be found in GenBank (Accession number MN881998.1). All the mitogenomes was submitted to GenBank with following accession numbers for each isolate used in this study: MW794295 for isolate B5-A4, MW794296 for isolate BG-B1-A17, MW794297 for isolate SC-B2-A4, MW794298 for T-B1-A5, MW794299 for isolate Ti-B3-A2, MW794300 for isolate Ti-B3-A3-2, MW794301 for isolate Yolkenari-1, MW794302 for isolate Yildirim-2-10th, MW794303 for isolate Yildirim-1, MW794304 for isolate T-B1-A4-2, MW794305 for isolate MT-B1-A3-1, MW794306 for isolate MM-B4-A4, MW794307 for isolate MM-B2-A2, and MW794308 for isolate 2B1-A5.

AUTHOR CONTRIBUTIONS

GY performed laboratory work and all the data analyses. HO designed the research, recommended, and directed the data analysis methods, and controlled all steps of the study. Both authors confirmed and discussed the results and wrote the article together.

FUNDING

This study was supported by TUBITAK (Scientific and Technological Research Council of Turkey) project no. 217Z134, granted to HO.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.647989/full#supplementary-material

Supplementary Table 1 | Characterization of annotated mitochondrial genes in *Monilinia fructicola*-Isolate BG-B1-A4, accession number: MT005827.

Supplementary Table 2 | Set of 32 tRNAs encoding different anticodons in the mitogenomes of *Monilinia fructicola* and *Monilinia laxa* species.

Supplementary Table 3 | Dataset of the relative synonymous codon usage (RSCU) and the total number of usages for each of the codons in the 14 mitochondrial PCGs, and rps3.

REFERENCES

Aguileta, G., De Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., et al. (2014). High variability of mitochondrial gene order among fungi. *Genome Biol. Evol.* 6, 451–465. doi: 10.1093/gbe/evu028

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham: Babraham Institute.

Badet, T., and Croll, D. (2020). The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr. Opin. Plant Biol.* 56, 65–73. doi: 10.1016/j.pbi.2020.04.009

Badet, T., Oggenfuss, U., Abraham, L., McDonald, B. A., and Croll, D. (2020). A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol.* 18:12. doi: 10.1186/s12915-020-0744-3

Badotti, F., de Oliveira, F. S., Garcia, C. F., Vaz, A. B. M., Fonseca, P. L. C., Nahum, L. A., et al. (2017). Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of *Basidiomycota* (Fungi). *BMC Microbiol.* 17:42. doi: 10.1186/s12866-017-0958-x

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Basse, C. W. (2010). Mitochondrial inheritance in fungi. *Curr. Opin. Microbiol.* 13, 712–719. doi: 10.1016/j.mib.2010.09.003

Beck, N., and Lang, B. (2010). *MFannot, Organelle Genome Annotation Websever*. Montréal QC: Université de Montréal.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., et al. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. doi: 10.1016/j.ympev.2012.08.023

Bertazzoni, S., Williams, A. H., Jones, D. A., Syme, R. A., Tan, K.-C., and Hane, J. K. (2018). Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic Fungi. *Mol. Plant Microbe Interact.* 31, 779–788. doi: 10.1094/MPMI-06-17-0135-FI

Blackwell, M. (2011). The Fungi: 1, 2, 3…5.1 million species? *Am. J. Bot.* 98, 426–438. doi: 10.3732/ajb.1000298

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brankovics, B., Kulik, T., Sawicki, J., Bilska, K., Zhang, H., de Hoog, G. S., et al. (2018). First steps towards mitochondrial pan-genomics: detailed analysis of *Fusarium graminearum* mitogenomes. *PeerJ* 6:e5963. doi: 10.7717/peerj.5963

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

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716. doi: 10.1016/j.tig.2003.10.012

Chen, C., Li, Q., Fu, R., Wang, J., Xiong, C., Fan, Z., et al. (2019). Characterization of the mitochondrial genome of the pathogenic fungus *Scytalidium auriculariicola* (Leotiomycetes) and insights into its phylogenetics. *Sci. Rep.* 9:17447. doi: 10.1038/s41598-019-53941-5

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147

Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* 35, 57–65. doi: 10.1016/j.gde.2015.09.001

Franco, M. E. E., López, S. M. Y., Medina, R., Lucentini, C. G., Troncozo, M. I., Pastorino, G. N., et al. (2017). The mitochondrial genome of the plant-pathogenic fungus *Stemphylium lycopersici* uncovers a dynamic structure due to repetitive and mobile elements. *PLoS One* 12:e0185545. doi: 10.1371/journal.pone.0185545

Fujimori, S., Washio, T., and Tomita, M. (2005). GC-compositional strand bias around transcription start sites in plants and fungi. *BMC Genomics* 6:26. doi: 10.1186/1471-2164-6-26

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Holb, I. (2006). Possibilities of brown rot management in organic stone fruit production in Hungary. *Int. J. Hortic. Sci.* 12, 87–91.

Holb, I. (2008). Brown rot blossom blight of pome and stone fruits: symptom, disease cycle, host resistance, and biological control. *Int. J. Hortic. Sci.* 14, 15–21.

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C. W., Yi, T.-S., et al. (2019). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *bioRxiv* [Preprint]. doi: 10.1186/s13059-020-02154-5

Joardar, V., Abrams, N. F., Hostetler, J., Paukstelis, P. J., Pakala, S., Pakala, S. B., et al. (2012). Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genome size variability. *BMC Genomics* 13:698. doi: 10.1186/1471-2164-13-698

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kelly, A. C., and Ward, T. J. (2018). Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. *PLoS One* 13:e0194616. doi: 10.1371/journal.pone.0194616

Kolesnikova, A. I., Putintseva, Y. A., Simonov, E. P., Biriukov, V. V., Oreshkova, N. V., Pavlov, I. N., et al. (2019). Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related *Armillaria* species. *BMC Genomics* 20:351. doi: 10.1186/s12864-019-5732-z

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

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Li, Q., Ren, Y., Xiang, D., Shi, X., Zhao, J., Peng, L., et al. (2020). Comparative mitogenome analysis of two ectomycorrhizal fungi (*Paxillus*) reveals gene rearrangement, intron dynamics, and phylogeny of basidiomycetes. *IMA Fungus* 11:12. doi: 10.1186/s43008-020-00038-8

Li, X., Li, L., Bao, Z., Tu, W., He, X., Zhang, B., et al. (2020). The 287,403 bp Mitochondrial Genome of Ectomycorrhizal Fungus *Tuber calosporum* Reveals Intron Expansion, tRNA Loss, and Gene Rearrangement. *Front. Microbiol.* 11:591453. doi: 10.3389/fmicb.2020.591453

Losada, L., Pakala, S. B., Fedorova, N. D., Joardar, V., Shabalina, S. A., Hostetler, J., et al. (2014). Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiol. Lett.* 352, 165–173. doi: 10.1111/1574-6968.12387

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413

Mardanov, A. V., Beletsky, A. V., Kadnikov, V. V., Ignatov, A. N., and Ravin, N. V. (2014). The 203 kbp mitochondrial genome of the phytopathogenic fungus Sclerotinia borealis reveals multiple invasions of introns and genomic duplications. *PLoS One* 9:e107536. doi: 10.1371/journal.pone.0107536

Marsolier-Kergoat, M.-C. (2013). Models for the evolution of GC content in asexual fungi *Candida albicans* and *C. dubliniensis*. *Genome Biol. Evol.* 5, 2205–2216. doi: 10.1093/gbe/evt170

McCarthy, C. G., and Fitzpatrick, D. A. (2019). Pan-genome analyses of model fungal species. *Microb. Genom.* 5:e000243. doi: 10.1099/mgen.0.000243

Medina, R., Franco, M. E. E., Bartel, L. C., Martinez Alcantara, V., Saparrat, M. C., and Balatti, P. A. (2020). Fungal mitogenomes: relevant features to planning plant disease management. *Front. Microbiol.* 11:978. doi: 10.3389/fmicb.2020.00978

Megarioti, A. H., and Kouvelis, V. N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.* 12, 1337–1354. doi: 10.1093/gbe/evaa126

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., and Bernardi, G. (2006). Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem. Biophys. Res. Commun.* 347, 1–3. doi: 10.1016/j.bbrc.2006.06.054

Nardi, F., Carapelli, A., and Frati, F. (2012). Repeated regions in mitochondrial genomes: distribution, origin and evolutionary significance. *Mitochondrion* 12, 483–491. doi: 10.1016/j.mito.2012.07.105

Nie, Y., Wang, L., Cai, Y., Tao, W., Zhang, Y. J., and Huang, B. (2019). Mitochondrial genome of the entomophthoroid fungus *Conidiobolus heterosporus* provides insights into evolution of basal fungi. *Appl. Microbiol. Biotechnol.* 103, 1379–1391. doi: 10.1007/s00253-018-9549-5

Nosek, J., Novotna, M., Hlavatovicova, Z., Ussery, D., Fajkus, J., and Tomaska, L. (2004). Complete DNA sequence of the linear mitochondrial genome of the pathogenic yeast *Candida parapsilosis*. *Mol. Genet. Genomics* 272, 173–180. doi: 10.1007/s00438-004-1046-0

Novoa, E. M., and de Pouplana, L. R. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28, 574–581. doi: 10.1016/j.tig.2012.07.006

Ozer, E. A., Allen, J. P., and Hauser, A. R. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* 15:737. doi: 10.1186/1471-2164-15-737

Ozkilinc, H., Yildiz, G., Silan, E., Arslan, K., Guven, H., Altinok, H. H., et al. (2020). Species diversity, mating type assays and aggressiveness patterns of *Monilinia* pathogens causing brown rot of peach fruit in Turkey. *Eur. J. Plant Pathol.* 157, 799–814. doi: 10.1007/s10658-020-02040-7

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Res.* 28, 395–404. doi: 10.1038/s41422-018-0013-y

Plissonneau, C., Hartmann, F. E., and Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* 16:5. doi: 10.1186/s12915-017-0457-4

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., et al. (2010). Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330, 1540–1543. doi: 10.1126/science.1193070

Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430. doi: 10.1038/nrmicro2790

Rambaut, A. (2012). *FigTree v1. 4*. Oxford: University of Oxford.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

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., Vicario, S., Pappadà, G., Scioscia, G., Scazzocchio, C., and Saccone, C. (2009). Towards barcode markers in Fungi: an intron map of Ascomycota mitochondria. *BMC Bioinformatics* 10(Suppl. 6):S15. doi: 10.1186/1471-2105-10-S6-S15

Sethuraman, J., Majer, A., Iranpour, M., and Hausner, G. (2009). Molecular evolution of the mtDNA encoded *rps3* gene among filamentous ascomycetes fungi with an emphasis on the Ophiostomatoid fungi. *J. Mol. Evol.* 69, 372–385. doi: 10.1007/s00239-009-9291-9

Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* 12, 335–337. doi: 10.1007/s13127-011-0056-0

Smits, P., Smeitink, J. A., van den Heuvel, L. P., Huynen, M. A., and Ettema, T. J. (2007). Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res.* 35, 4686–4703. doi: 10.1093/nar/gkm441

Stajich, J. E. (2017). "Fungal genomes and insights into the evolution of the kingdom," in *The* Microbiol. Spectr. 5:10.1128/microbiolspec.FUNK-0055-2016.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102–1104. doi: 10.2144/00286ir01

Stukenbrock, E. H., and Dutheil, J. Y. (2018). Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics* 208, 1209–1229. doi: 10.1534/genetics.117.300502

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006

Torres, D. E., Oggenfuss, U., Croll, D., and Seidl, M. F. (2020). Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol. Rev.* 34, 136–143. doi: 10.1016/j.fbr.2020.07.001

Torriani, S. F., Penselin, D., Knogge, W., Felder, M., Taudien, S., Platzer, M., et al. (2014). Comparative analysis of mitochondrial genomes from closely related *Rhynchosporium* species reveals extensive intron invasion. *Fungal Genet. Biol.* 62, 34–42. doi: 10.1016/j.fgb.2013.11.001

Wei, L., He, J., Jia, X., Qi, Q., Liang, Z., Zheng, H., et al. (2014). Analysis of codon usage bias of mitochondrial genome in *Bombyx moriand* its relation to evolution. *BMC Evol. Biol.* 14:262. doi: 10.1186/s12862-014-0262-4

Yildiz, G., and Ozkilinc, H. (2020). First characterization of the complete mitochondrial genome of fungal plant-pathogen *Monilinia laxa* which represents the mobile intron rich structure. *Sci. Rep.* 10:13644. doi: 10.1038/s41598-020-70611-z

Conflict of Interest: The 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fmicb.2021.646567

Characterization of the Complete Mitochondrial Genome of Basidiomycete Yeast *Hannaella oryzae*: Intron Evolution, Gene Rearrangement, and Its Phylogeny

Qiang Li[1], Lijiao Li[1], Huiyu Feng[1], Wenying Tu[1], Zhijie Bao[1], Chuan Xiong[2], Xu Wang[3], Yuan Qing[4]* and Wenli Huang[2]*†*

[1] Key Laboratory of Coarse Cereal Processing, Ministry of Agriculture and Rural Affairs, School of Food and Biological Engineering, Chengdu University, Chengdu, China, [2] Biotechnology and Nuclear Technology Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu, China, [3] College of Life Sciences, Henan Agricultural University, Zhengzhou, China, [4] Panxi Featured Crops Research and Utilization Key Laboratory of Sichuan Province, Xichang University, Xichang, China

In this study, the mitogenome of *Hannaella oryzae* was sequenced by next-generation sequencing (NGS) and successfully assembled. The *H. oryzae* mitogenome comprised circular DNA molecules with a total size of 26,444 bp. We found that the mitogenome of *H. oryzae* partially deleted the tRNA gene transferring cysteine. Comparative mitogenomic analyses showed that intronic regions were the main factors contributing to the size variations of mitogenomes in Tremellales. Introns of the *cox1* gene in Tremellales species were found to have undergone intron loss/gain events, and introns of the *H. oryzae cox1* gene may have different origins. Gene arrangement analysis revealed that *H. oryzae* contained a unique gene order different from other Tremellales species. Phylogenetic analysis based on a combined mitochondrial gene set resulted in identical and well-supported topologies, wherein *H. oryzae* was closely related to *Tremella fuciformis*. This study represents the first report of mitogenome for the *Hannaella* genus, which will allow further study of the population genetics, taxonomy, and evolutionary biology of this important phylloplane yeast and other related species.

Keywords: *Hannaella*, mitochondrial genome, protein coding gene, repeat sequence, gene rearrangement, phylogenetic analysis 3

OPEN ACCESS

Edited by:
Georg Hausner,
University of Manitoba, Canada

Reviewed by:
Sajeet Haridas,
Joint Genome Institute, United States
Yongjie Zhang,
Shanxi University, China

****Correspondence:***
Xu Wang
xuwang@henau.edu.cn
Yuan Qing
64899867@qq.com
Wenli Huang
wenlih11@126.com

†Present address:
Wenli Huang,
Sichuan Academy of Agricultural Sciences, Chengdu, China

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *27 December 2020*
Accepted: *19 April 2021*
Published: *28 May 2021*

Citation:
Li Q, Li L, Feng H, Tu W, Bao Z, Xiong C, Wang X, Qing Y and Huang W (2021) Characterization of the Complete Mitochondrial Genome of Basidiomycete Yeast Hannaella oryzae: Intron Evolution, Gene Rearrangement, and Its Phylogeny. Front. Microbiol. 12:646567. doi: 10.3389/fmicb.2021.646567

INTRODUCTION

Hannaella is a basidiomycetous yeast genus belonging to the order Tremellales, phylum Basidiomycota. It was proposed to accommodate species closely related to the genera *Derxomyces* and *Dioszegia*, which belong to the *Bullera sinensis* clade of the *Luteolus* lineage of the Tremellales based on a series of molecular markers (Wang and Bai, 2008; Kaewwichian et al., 2015; Han et al., 2017). So far, about 12 species have been described in this genus, including *Hannaella*

coprosmaensis, *Hannaella dianchiensis*, *Hannaella kunmingensis*, *Hannaella luteola*, *Hannaella phyllophila*, *Hannaella phetchabunensis*, *Hannaella pagnoccae*, *Hannaella surugaensis*, *Hannaella sinensis*, *Hannaella siamensis*, *Hannaella zeae*, and *Hannaella oryzae* (Landell et al., 2014; Kaewwichian et al., 2015; Han et al., 2017). These species are found widely distributed on the leaf surfaces of various plants, including rice, wheat, and fruit trees (Han et al., 2017). As an important phyllosphere-inhabiting yeast, *H. oryzae* was considered to play an important role in promoting plant growth and biocontrol. This genus is considered to be monophyletic in nature, and *H. kunmingensis* showed genotypic and phenotypic variability (Dayo-Owoyemi et al., 2013). Both the basidiomycetous and the ascomycetous yeasts have been found colonizing on phylloplane from various regions of the world (Glushakova et al., 2007; Landell et al., 2010; Molnarova et al., 2014). These basidiomycetous genera were found to be the most common phylloplane yeasts, including *Sporobolomyces*, *Rhodotorula*, *Cryptococcus*, *Trichosporon*, and *Hannaella* (de Azeredo et al., 1998; Kaewwichian et al., 2015). The mitochondrial genomic characteristics of representative species of the two phylloplane yeast genera (*Cryptococcus* and *Trichosporon*) have been published, which facilitated our understanding of phylloplane yeasts (Yang et al., 2012; Gomes et al., 2018; Yan et al., 2018). However, no mitochondrial genome has been published on the genus *Hannaella* or the family Bulleribasidiaceae. *H. oryzae* has been isolated from the phylloplane of various plants. The report of its mitochondrial genome will help us understand the genetic characteristics of this important phylloplane yeast.

Mitochondrial genomes are effective tools for analyzing the phylogenetic and genetic evolution of eukaryotes because they contain many available molecular markers (Burger et al., 2003; Bullerwell and Lang, 2005; Cheng et al., 2021). Besides, the arrangement of mitochondrial genes, their transfer RNA (tRNA) structure, their codon usage, and the dynamic changes of introns can also be used to infer the evolutionary status of eukaryotes (Li et al., 2018a; Wang et al., 2018, 2020b; Wu et al., 2021). With the rapid development of next-generation sequencing technologies in recent years, more and more mitochondrial genomes have been obtained (du Toit et al., 2017; Yang et al., 2018; Zhang et al., 2018; Wang et al., 2020c). However, the mitochondrial genomes of fungi are less studied than those of animals. So far, less than 130 basidiomycete mitochondrial genomes have been reported[1], which limits our understanding of the "second genome" (mitochondrial genome) of fungi. Studies on the fungal mitogenomes have shown that fungal mitogenomes exhibited significant variations in gene order, introns, intergenic regions, genome size, and repetitive sequences (Zhang et al., 2016; Zhang S. et al., 2017; Zhang Y.J. et al., 2017; Fourie et al., 2018; Wang et al., 2020a; Li et al., 2021). Despite enormous variations in the fungal genome, the 15 protein-coding genes, including *atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, and *rps3*, have been detected in most basidiomycete mitochondrial genomes, which were considered to be core protein-coding genes (PCGs) in the basidiomycete mitochondrial genomes.

In the present study, the complete mitochondrial genome of *H. oryzae* was sequenced and assembled by next-generation sequencing technology. The content, organization, and structure of the mitochondrial genes were revealed. We compared the mitochondrial genome of *H. oryzae* with its closely related species to identify variations and similarities in the gene content, genome organization, and gene order. The dynamic changes of introns were also revealed in *H. oryzae* and other basidiomycete species. In addition, the phylogenetic relationships among various basidiomycete species based on combined mitochondrial gene sets were analyzed. The mitochondrial genome of *H. oryzae* will allow further study of the population genetics, taxonomy, and evolutionary biology of this important phylloplane yeast and other related species.

MATERIALS AND METHODS

Sample Collection, DNA Extraction, and Sequencing

The *H. oryzae* strain s11 was isolated from the surface of corn leaves collected in Sichuan, China, using the improved ballistoconidia-fall method as described by Kaewwichian et al. (2015). The morphological, biochemical, and physiological characteristics of the collected yeast strains were examined according to standard methods described by Kurtzman et al. (2011). The total genomic DNA was extracted using the method described by Wang and Bai (2008). *H. oryzae* was further identified based on the internal transcribed spacer (ITS) sequence and the mitochondrial *cob* gene. Whole-genome sequencing libraries were constructed using NEBNext Ultra II DNA Library Prep Kits (NEB, Beijing, China) with the extracted genomic DNA following the manufacturer's instructions. Whole-genome sequencing was carried out on an Illumina HiSeq 2500 Platform (Illumina, San Diego, CA, United States). To verify the accuracy and integrity of our assembled genome, we further sequenced the genomic DNA using the Pacbio RSII platform (Pacific Biosciences, CA, United States). A 40-kb SMRTbell DNA library was prepared to perform the Pacbio sequencing.

De novo Assembly and Annotation of the *H. oryzae* Mitogenome

Illumina PCR adapter reads and low-quality reads from the paired-end reads were filtered using custom scripts. Clean reads were obtained after the quality control step. The mitogenome of *H. oryzae* was assembled by CANU v1.6 (Koren et al., 2017) using the Pacbio long reads. The assembled contigs were further polished using the paired-end Illumina reads with Pilon v1.22 (Walker et al., 2014). The obtained *H. oryzae* complete mitogenome was annotated according to the methods previously described (Li et al., 2020b,c). Briefly, the complete mitogenome of *H. oryzae* was firstly annotated based on the results of MITOS (Bernt et al., 2013) and MFannot

[1]https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/

(Valach et al., 2014). At this step, PCGs, ribosomal RNA (rRNA) genes, and tRNA genes were initially annotated. Open reading frames (ORFs) were modified or predicted with the NCBI Open Reading Frame Finder [ORFs less than 100 amino acids (aa) were excluded] (Coordinators, 2017) and annotated with BLASTp searches against the NCBI non-redundant protein sequence database (Bleasby and Wootton, 1990). tRNA genes were also predicted with tRNAscan-SE v1.3.1 (Lowe and Chan, 2016). The graphical map of the complete mitogenome was drawn with OGDraw v1.2 (Lohse et al., 2013).

Sequence Analysis

The base composition of the *H. oryzae* mitogenome was analyzed using DNASTAR Lasergene v7.1[2] software. Strand asymmetry of the *H. oryzae* mitogenome was assessed using the following formulas: AT skew = [A − T]/[A + T] and GC skew = [G − C]/[G + C] (Wang et al., 2017). The non-synonymous substitution rate (Ka) and the synonymous substitution rate (Ks) for the core PCGs in the four Tremellales mitogenomes were calculated using DnaSP v6.10.01 (Rozas et al., 2017). We used MEGA v6.06 (Caspermeyer, 2016) to calculate the overall mean genetic distances between each pair of the 15 core PCGs (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, and *rps3*) using the Kimura-2-parameter (K2P) model. The genome synteny of the closely related mitogenomes was analyzed using Mauve v2.4.0 (Darling et al., 2004). The introns of the *cox1* gene in 33 basidiomycete species were classified into different position classes (Pcls) and named according to previously described methods (Zhang and Zhang, 2019).

Phylogenetic Analysis

In order to investigate the phylogenetic status of *H. oryzae* among the Basidiomycota phylum, we constructed a phylogenetic tree of 33 species based on the combined mitochondrial gene set (15 core PCGs + two rRNA genes) (Li et al., 2018d). The individual mitochondrial gene was first aligned using MAFFT v7.037 (Katoh et al., 2019). Then, these alignments were concatenated in SequenceMatrix v1.7.8 (Vaidya et al., 2011). In order to detect potential phylogenetic conflicts between different genes, we carried out a preliminary partition homogeneity test. Phylogenetic trees were constructed using both Bayesian inference (BI) and maximum likelihood (ML) methods. Best-fit models of evolution and partitioning schemes for the gene set were determined according to PartitionFinder 2.1.1 (Lanfear et al., 2017). BI analysis was performed with MrBayes v3.2.6 (Ronquist et al., 2012). The RAxML v 8.0.0 software (Stamatakis, 2014) was used for ML analysis.

Data Availability Statement

The newly sequenced *H. oryzae* mitogenome was deposited in the GenBank database under the following Accession No.: MH732752.

[2]http://www.dnastar.com/

RESULTS

Protein-Coding Genes, RNA Genes, and Codon Analysis in the *H. oryzae* Mitogenome

The complete mitochondrial genome of *H. oryzae* was assembled into a circular DNA molecule with a total size of 26,444 bp (**Figure 1**). The GC content of the *H. oryzae* mitogenome was 38.98% (**Supplementary Table 1**). Both the GC skew and the AT skew of the *H. oryzae* mitochondrial genome were positive. Fifteen conserved PCGs were detected in the *H. oryzae* mitogenome, including 14 core PCGs for energy metabolism and one *rps3* gene (**Supplementary Table 2**). Eight introns were detected in the mitogenome of *H. oryzae* distributed in seven host genes, including *cox1*, *atp6*, *rnl*, *atp9*, *nad3*, *nad1*, and *nad5*. The *cox1* gene contained two introns, while the other six genes each contained one intron. All these introns belonged to group I, which could catalyze their own splicing. One intronic ORF (*orf195*) was found in the *H. oryzae* mitochondrial genome, which encoded LAGLIDADG homing endonuclease. Of the 42 genes detected in the *H. oryzae* mitogenome, 19 were on the direct strand and the other 23 were on the reverse strand.

The mitogenome of *H. oryzae* contained two rRNA genes, namely the small subunit ribosomal RNA (*rns*) and the large subunit ribosomal RNA (*rnl*) (**Supplementary Table 2**). Twenty-three tRNA genes were detected in the *H. oryzae* mitogenome, encoding for 19 standard amino acids. The *H. oryzae* mitogenome lacked a tRNA gene that encoded for amino acid cysteine. Interestingly, we found that all four species from the order Tremellales lacked the tRNA gene to transfer cysteine. We detected about 40 bases of homology to Cys-tRNAs in the four mitogenomes. In addition, the *H. oryzae* mitogenome contained two tRNAs with the same anticodons coding for methionine and proline (**Supplementary Figure 1**). However, these two tRNAs varied in length and base composition, indicating frequent gene mutations in tRNA genes. The *H. oryzae* mitogenome also contained two tRNAs with different anticodons that coded for leucine and serine. The mitogenome of *H. oryzae* lost a tRNA gene with different anticodons that coded for arginine. The length of individual tRNAs ranged from 71 to 85 bp, which was mainly due to the variations in length of the extra arm. A ribonuclease P RNA gene (*rnpB*) was found in the *H. oryzae* mitogenome with a length of 211 bp.

The sizes of the five mitogenomes varied greatly, ranging from 24,874 to 35,058 bp. Correlation analysis showed that the intronic regions had the greatest impact on the size variations of Tremellales and Trichosporonales mitogenomes, with Pearson's correlation coefficient of 0.907 ($P < 0.05$; **Supplementary Figure 2**). The results suggested that the intronic region was the main factor promoting the size variations of mitogenomes in Tremellales and Trichosporonales.

Intron Evolution of *H. oryzae* Mitogenome

Introns were considered as the main factors leading to variations of the mitogenome size in Tremellales, Trichosporonales, and

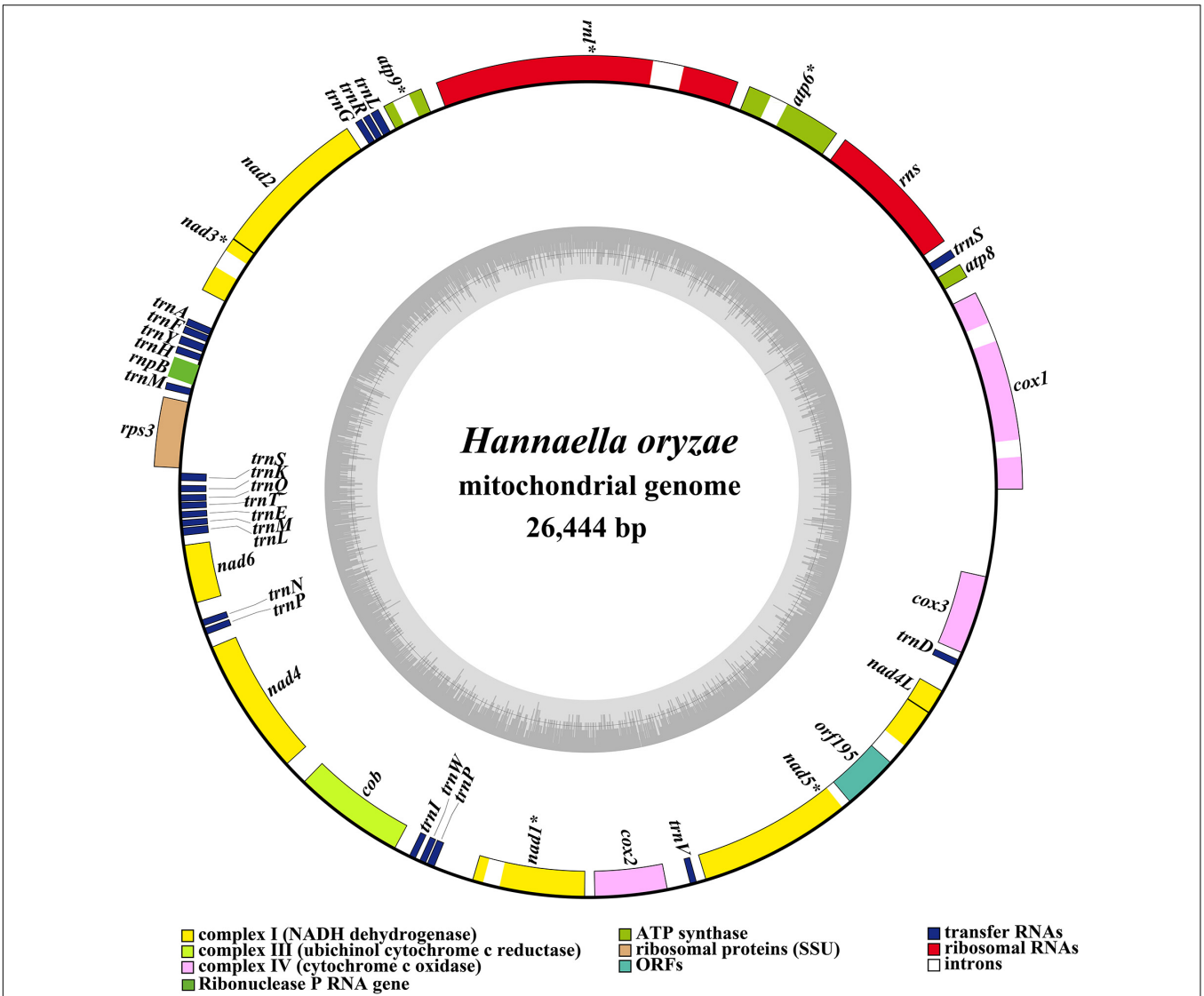


FIGURE 1 | Circular map of the mitochondrial genome of *Hannaella oryzae*. Genes are represented by *different colored blocks*. *Colored blocks outside each ring* indicate that the genes are on the direct strand, while *colored blocks within the ring* indicate that the genes are located on the reverse strand. The *inner grayscale bar graph* shows the GC content of the mitochondrial sequences. The *circle inside the GC content graph* marks the 50% threshold. The graphical map of the complete mitogenome was drawn with OGDraw v1.2 (Lohse et al., 2013).

other basidiomycetes. So the dynamic changes of the introns in *H. oryzae* and other basidiomycetes were analyzed in depth in the present study. The *cox1* gene was found to be the largest host gene of introns in basidiomycetes (Ferandon et al., 2010). Introns in the *cox1* gene could be divided into different Pcls according to the insertion sites in the coding region. The introns from the same Pcls were considered homologous introns with high sequence similarities (Ferandon et al., 2010). Introns belonging to different Pcls were considered to be non-homologous and contained low sequence similarities. In the present study, a total of 246 introns of the *cox1* gene were detected in the 33 basidiomycete species we tested, seven of which belonged to group II and the others belonged to group I (**Figure 2**). The 239 group I introns could be classified into 31 Pcls, 26 of which were reported in previous studies (Ferandon et al., 2010), while the other five were novel Pcls found in the 33 basidiomycetes. Among these Pcls, Pcl P386 was the most widely distributed, distributed in 21 of 33 basidiomycete species. Pcls P615 and P709 were also common Pcls in basidiomycetes, both of which existed in 16 of the 33 basidiomycete species. Pcls P196, P221, and P941 were found only in one of the 33 basidiomycete species, which were considered as rare introns in basidiomycetes. Eight introns were detected in the *cox1* gene of five species from Tremellales and Trichosporonales. *Cryptococcus neoformans* and *Tremella fuciformis* lost introns in the *cox1* gene, while the *cox1* gene of *Cryptococcus gattii* contained the most introns among Tremellales and Trichosporonales, indicating that species in Tremellales and Trichosporonales had experienced intron

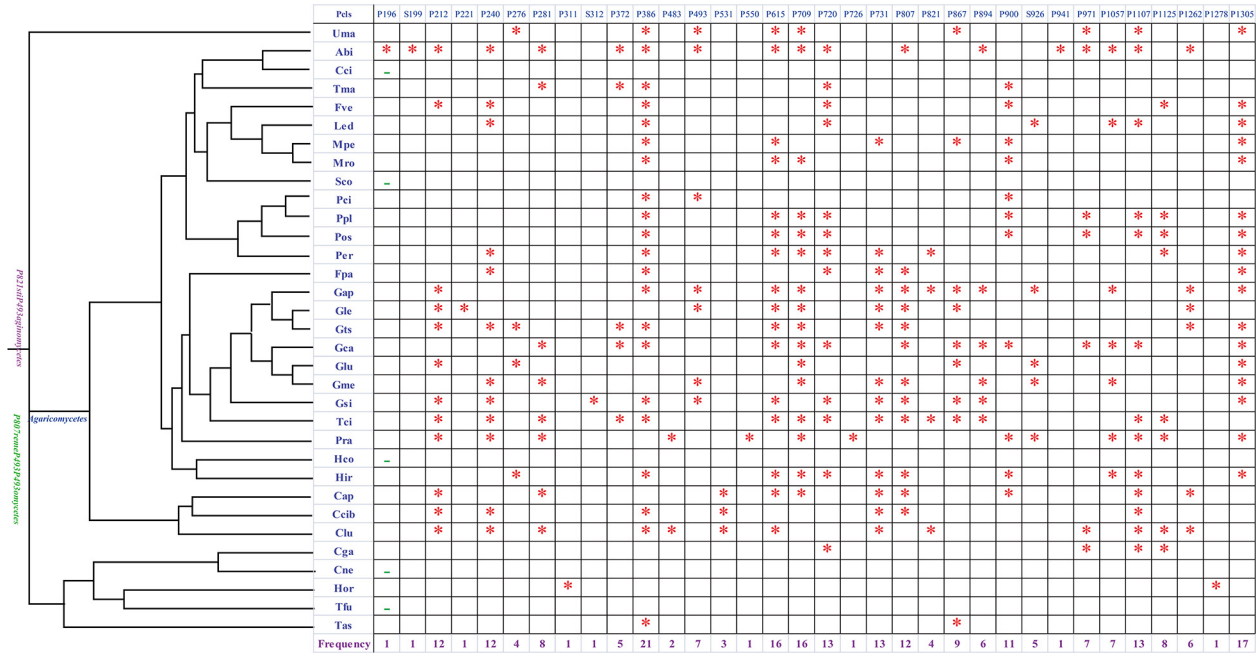


FIGURE 2 | Position class (Pcl) information of *cox1* introns in 33 basidiomycete species. The phylogenetic tree on the *left* was constructed using the Bayesian inference (BI) and maximum likelihood (ML) methods based on a combined mitochondrial gene set (Bayesian posterior probabilities and bootstrap support values are shown in **Figure 5**). Pcls are named according to previously described methods (Zhang and Zhang, 2019). Species IDs for the mitogenomes used in the intron analysis are provided in **Supplementary Table 4**. *Asterisk* indicates the presence of the Pcl in the species and an *en dash* indicates the absence of any Pcl in the species. The *number at the bottom* indicates the frequency of occurrence of the Pcl in the 33 basidiomycete species.

loss/gain events during evolution. In addition, *H. oryzae* was found to contain two novel intron Pcls at sites 103 and 425 aa (**Supplementary Table 3**). No homologous introns were found in the four closely related species and other basidiomycete species, indicating that *H. oryzae* had undergone a unique intron origin.

Variations, Genetic Distance, and Evolution Rate of Core Genes

Of the 15 core PCGs detected in the four Tremellales mitogenomes, 12 PCGs were found to vary in length among the four species, except for *atp8*, *atp9*, and *nad4L* (**Supplementary Figure 3**). Nine genes of the *H. oryzae* mitogenome had unique lengths among the four Tremellales mitogenomes, including the *cob*, *cox1*, *cox3*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*, and *rps3* genes. The GC contents of the core PCGs in the four Tremellales mitogenomes also varied, indicating frequent base variations in the core PCGs of the Tremellales mitogenomes. Among the 15 core PCGs detected, *atp9* contained the highest GC content in all four Tremellales mitogenomes. The AT and GC skews of most of the core PCGs were negative in the four mitogenomes. The GC skews of the 15 core PCGs in the *H. oryzae* mitogenome were negative. The *rps3* gene contained a positive AT skew in the *H. oryzae* mitogenome.

Of the 15 core PCGs detected, the *rps3* gene had the highest mean K2P genetic distance among the four Tremellales mitogenomes, followed by *nad3* (**Supplementary Figure 4**). The mean genetic distance of *atp9* in the four mitogenomes was the smallest among the 15 core PCGs, indicating that this gene was highly conserved across the mitogenomes. *nad4L* contained the highest mean Ks among the 15 core PCGs detected, while *atp9* had the lowest rate. The highest Ka was observed in the *rps3* gene, while *nad4L* exhibited the lowest Ka value among the 15 PCGs detected. The Ka/Ks values for all 15 PCGs were less than 1, indicating that these genes were subject to purifying selection.

Comparative Genome Analysis

The mitogenome size of *H. oryzae* was the second smallest among all published Basidiomycota mitochondrial genomes (see text footnote 1), only higher than that of the basidiomycete yeast *C. neoformans* (Yan et al., 2018; **Supplementary Table 1**). The mitogenome size of *H. oryzae* was smaller than that of mushroom-forming species, such as *Pleurotus* spp. (Li et al., 2018a), *Ganoderma* spp. (Li et al., 2018d, 2019c), *Coprinopsis cinereal* (Stajich et al., 2010), and *Agaricus bisporus* (Ferandon et al., 2013), and its closely related species, *T. fuciformis* (MF422647); smaller than that of the ectomycorrhizal fungi *Tricholoma matsutake* (Yoon et al., 2016), *Lactarius* spp. (Li et al., 2019b), *Russula* spp.(Li et al., 2018c), and *Cantharellus* spp. (Li et al., 2018b); and also smaller than that of the basidiomycete yeast *C. gattii* (Yadav et al., 2018). The GC content of the *H. oryzae* mitogenome was relatively high (38.98%), which was only lower than that of *Rhodotorula mucilaginosa* (40.43%) (Gan et al., 2017), among all the published Basidiomycota mitochondrial genomes detected. There were seven, seven, two, and four

FIGURE 3 | Comparison of the conserved gene order among the mitochondrial genomes of the five closely related species in Tremellales and Trichosporonales, including *Cryptococcus gattii*, *Cryptococcus neoformans*, *Hannaella oryzae*, *Tremella fuciformis*, and *Trichosporon asahii*. Fifteen core protein-coding genes (PCGs) and two rRNA genes were included in this analysis. Genes are represented by *different colored blocks*.

introns detected in the mitogenomes of *T. fuciformis*, *C. gattii*, *C. neoformans*, and *Trichosporon asahii*, which contained five, six, one, and four introns, respectively. However, eight introns were detected in the mitogenome of *H. oryzae*, but only one of them contained intronic ORF, suggesting that the mitogenome introns of *H. oryzae* were undergoing constriction.

Gene Rearrangement Analysis

The gene arrangement in the mitogenomes of five closely related species in Tremellales and Trichosporonales was highly variable (**Figure 3**). Of the 18 genes detected in the five closely related species, including the 15 core PCGs, two rRNA genes, and one *rnpB* gene, the relative positions of the 16 genes varied among the five mitogenomes. The gene arrangements of *C. gattii* and *C. neoformans* were highly conserved between the two species. However, at the class level, the arrangement of the mitochondrial genes was highly variable. *H. oryzae* contained a unique gene arrangement in the order Tremellales, indicating that gene rearrangements have occurred during the evolution of *H. oryzae*, involving protein-encoded genes, rRNA genes, and the *rnpB* gene.

The genome collinearity analysis showed that the five mitogenomes of Tremellales and Trichosporonales could be divided into 18 homologous regions (**Figure 4**). The relative positions of these homologous regions were highly variable among the five Tremellales mitogenomes. Out of the 18 homologous regions, the relative positions of 17 homologous regions varied among the five mitogenomes. The relative positions of the homologous regions were identical between *C. gattii* and *C. neoformans*.

Phylogenetic Analysis

We obtained identical and well-supported tree topologies using both BI and ML methods based on the combined mitochondrial

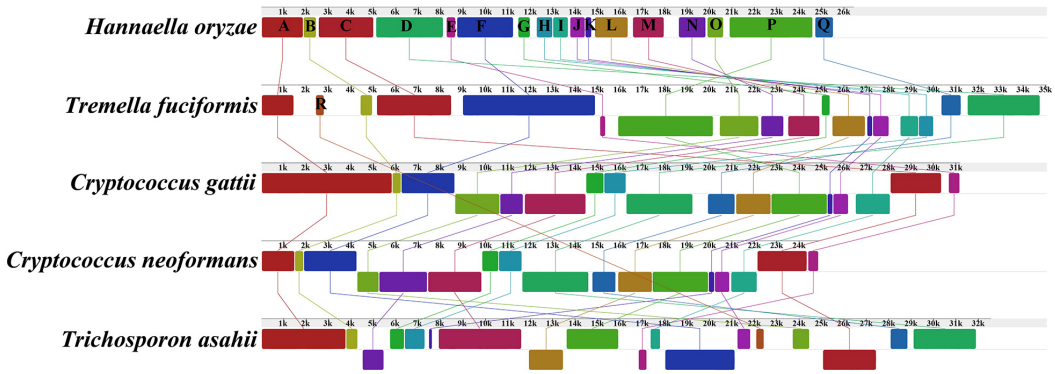


FIGURE 4 | Collinearity analysis of the mitochondrial genomes of the five closely related species in Tremellales and Trichosporonales, including *Cryptococcus gattii*, *Cryptococcus neoformans*, *Hannaella oryzae*, *Tremella fuciformis*, and *Trichosporon asahii*. Eighteen homologous regions were detected among the five mitogenomes. The sizes and the relative positions of the homologous regions varied among the mitogenomes. The genome synteny of the closely related mitogenomes was analyzed using Mauve v2.4.0 (Darling et al., 2004).

gene set (15 core PCGs + two rRNA genes) (**Figure 5**). All major clades of the trees were well supported [Bayesian posterior probability (BPP) = 1.00, bootstrap (BS) = 100]. Based on the phylogenetic analyses, the 33 Basidiomycota species could be divided into seven major clades corresponding to the orders Ustilaginales, Agaricales, Polyporales, Russulales, Cantharellales, Tremellales, and Trichosporonales (**Supplementary Table 4**). The four species from Tremellales could be divided into two groups: one group was composed of two species in the *Cryptococcus* genus and the second group was composed of *H. oryzae* and *T. fuciformis*. *H. oryzae* was identified as a sister species to *T. fuciformis*. The results showed that the combined mitochondrial gene set was suitable as a reliable molecular marker for the analysis of the phylogenetic relationships among Basidiomycota species.

DISCUSSION

As the "second genome" of eukaryotes, the mitochondrial genome plays an important role in regulating the growth and development, energy metabolism, aging, and stress resistance of eukaryotes (Bhargava and Schnellmann, 2017; Scheede-Bergdahl and Bergdahl, 2017). With the development of the next-generation sequencing technology, more and more mitochondrial genomes have been obtained, which has promoted our understanding of the origin, evolution, and taxonomy of eukaryotes (du Toit et al., 2017; Yang et al., 2018; Zhang et al., 2018). However, compared with animals, the mitochondrial genomes of fungi were less studied, especially that of Basidiomycota. It is estimated that there are more than 30,000 species of Basidiomycota in nature. However, up to now, less than 100 complete mitochondrial genomes of Basidiomycota are available in the public database (see text footnote 1). Only four mitochondrial genomes of basidiomycete yeasts were reported. This report of the mitochondrial genome of the important phylloplane yeast *H. oryzae* will broaden our understanding of basidiomycete mitochondrial genomes. The mitochondrial genome size of *H. oryzae* is the second smallest among all published Basidiomycota mitochondrial genomes, smaller than those of many mushroom-forming fungi (Stajich et al., 2010; Yadav et al., 2018), ectomycorrhizal fungi (Li et al., 2018b, 2020c), and that of its related species, *T. fuciformis*. However, the GC content of the *H. oryzae* mitogenome is the second highest among all published Basidiomycota mitochondrial genomes, indicating its unique mitogenome characteristics. Comparative mitogenomic analysis showed that the intron region had the greatest effect on the variations of the mitochondrial genome size of Tremellales, followed by the intergenic region and the protein-coding region, which was consistent with previous studies (Li et al., 2018b). Introns of the Tremellales *cox1* gene were found to have undergone intron loss/gain events, which resulted in the size and organization variations of the Tremellales mitogenomes. In addition, the *cox1* gene of *H. oryzae* contained two novel introns, which showed different origins from the known introns in basidiomycetes. Interestingly, we did not find any intron homologous to the introns of the *H. oryzae cox1* gene in the NCBI database by BlastN search. More mitochondrial genomes need to be obtained to reveal the origin of the introns in *H. oryzae* or other basidiomycete species. Mitochondrial genes of *H. oryzae* were found on both strands as compared with the ascomycete mitochondrial genes, which are usually on the same strand (Aguileta et al., 2014).

It was reported that the mitochondrial genome of eukaryotes was obtained from a common alpha-proteobacterium ancestor (Lang et al., 1999). As evolution progresses, most mitochondrial genes were transferred to the nuclear genome, which was considered to have many advantages (Adams et al., 2002; Adams and Palmer, 2003). However, most fungi retain 14 conserved protein-coding genes for energy metabolism and one *rps3* gene, which were called the core PCGs of the fungal mitogenome (Ye et al., 2020). In this study, these core PCGs were found to have high variation rates in the length and base composition even among the closely related species. Nine genes of *H. oryzae* were found to have unique lengths in the five closely related species, indicating the unique evolutionary characteristics of *H. oryzae*.

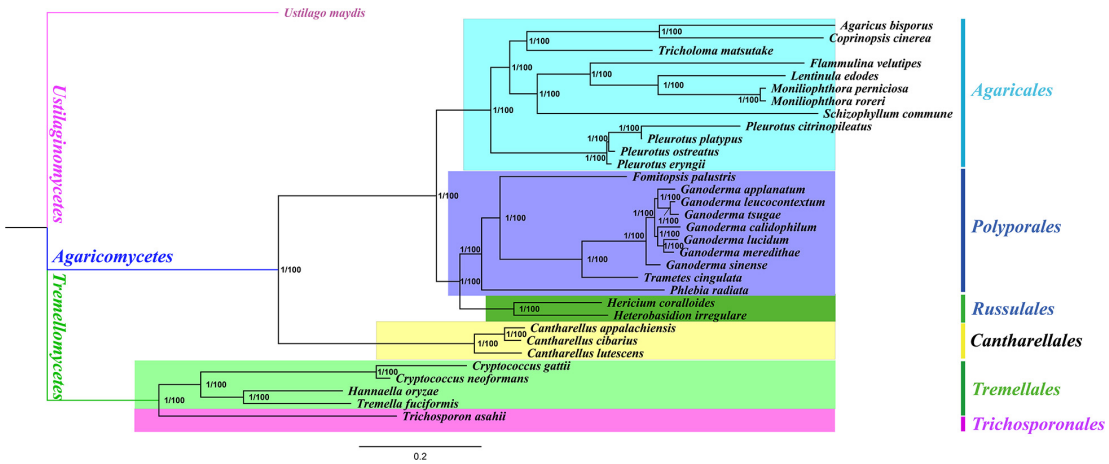


FIGURE 5 | Molecular phylogeny of 33 Basidiomycota species based on both Bayesian inference (BI) and maximum likelihood (ML) analyses of 15 core protein-coding genes and two rRNA genes. Support values are Bayesian posterior probabilities (BPP; *before the slash*) and bootstrap (BS; *after the slash*). The species and the NCBI accession numbers for the genomes used in the phylogenetic analysis are provided in **Supplementary Table 4**.

Of all core PCGs, *atp9* was found the most conserved among the five closely related species we studied. The genetic distance of the *rps3* gene was the largest among the five closely related species we examined, which was consistent with the report in other literature (Li et al., 2018b, 2020c). Interestingly, we found that the *atp8* gene in Tremellomycetes and Ustilaginomycetes had 12 nucleotides missing compared with that in Agaricomycetes species; the effect of the base variation of PCGs on the energy metabolism of species needs to be examined. In addition, we found that the core PCGs of the five closely related species in Tremellales and Trichosporonales were subject to purifying selection.

The arrangement of mitochondrial genes can serve as an important reference for revealing the evolutionary position of species (Li et al., 2018a; Zheng et al., 2018). Mitochondrial gene arrangement has been extensively studied in animals, and several models have been proposed to explain the rearrangement of mitochondrial genes in animals (Boore, 1999; Perseke et al., 2008). Compared with animals, the arrangement of fungal mitochondrial genes is highly variable (Hamari et al., 2001; Aguileta et al., 2014; Li et al., 2018a), and the gene rearrangement of the fungal mitochondrial genome has not been fully understood. Previous studies have shown that the accumulation of repetitive sequences in the fungal mitogenome has caused the fungal mitogenome rearrangement (Aguileta et al., 2014). In the present study, we found that the arrangement of mitochondrial genes was highly variable in the five closely related species, and the mitogenome of *H. oryzae* contained a unique gene order. In addition, the gene order varied at different taxonomic levels and is highly conserved between two species of the *Cryptococcus* genus. However, large-scale gene rearrangements have been observed at the class level, suggesting that Tremellales species undergo large-scale rearrangement events during evolution. Mitochondrial genes are also widely used as important molecular markers in the study of evolution, phylogeny, and population genetics (Bronstein and Kroh, 2018; Dai et al., 2018; Doyle et al., 2018; Li et al., 2020a). In this study, we divided 33 Basidiomycota species into seven clades based on a phylogenetic tree with high support for 15 core PCGs and two rRNA genes. Because there are few morphological features for fungi to be identified, this leads to confusion in fungal taxonomy and affects the phylogenetic research and utilization of fungi. The introduction of the mitochondrial genome will promote the understanding of fungal taxonomy and genetic evolution and can be used as a reliable tool for the analysis of fungal phylogeny (Li et al., 2018c, 2019a). More fungal mitochondrial genomes need to be uncovered to reconstruct the phylogenetic tree of fungi and to clarify the phylogenetic status of the Basidiomycota species.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

QL and WH conceived and designed experiments. QL, LL, HF, WT, ZB, CX, and XW analyzed the data. QL, YQ, and WH wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Outstanding Youth Fund of Sichuan Province (2019JDJQ0034), the Liangshan Science and Technology and Intellectual Property Office Project

(17NYCX0023), the Talent Fund of Sichuan Academy of Agricultural Sciences (2019QYXK003), and the High Tech Field Expansion Project of Sichuan Academy of Agricultural Sciences (2018GXTZ-001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.646567/full#supplementary-material

Supplementary Figure 1 | Putative secondary structures of the 23 tRNA genes identified in the mitogenome of *H. oryzae*. All genes are shown in order of occurrence in the mitogenome of *H. oryzae,* starting from *trnS*.

Supplementary Figure 2 | Pearson correlation analyses between mitochondrial genome sizes and different mitochondrial components of five closely related species in Tremellales and Trichosporonales.

Supplementary Figure 3 | Variation in the length and base composition of each of 15 core protein-coding genes (PCGs) among the mitochondrial genomes of the four closely related species in Tremellales. **(A)** PCG length variation; **(B)** GC content of the PCGs; **(C)** AT skew; **(D)** GC skew.

Supplementary Figure 4 | Genetic analysis of 15 core protein-coding genes conserved in the four closely related species in Tremellales. K2P, the Kimura-2-parameter distance; Ka, the mean number of non-synonymous substitutions per non-synonymous site; Ks, the mean number of synonymous substitutions per synonymous site.

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

Adams, K. L., Qiu, Y. L., Stoutemyer, M., and Palmer, J. D. (2002). Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9905–9912. doi: 10.1073/pnas.042694899

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., et al. (2014). High variability of mitochondrial gene order among fungi. *Genome Biol. Evol.* 6, 451–465. doi: 10.1093/gbe/evu028

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., et al. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. doi: 10.1016/j.ympev.2012.08.023

Bhargava, P., and Schnellmann, R. G. (2017). Mitochondrial energetics in the kidney. *Nat. Rev. Nephrol.* 13, 629–646. doi: 10.1038/nrneph.2017.107

Bleasby, A. J., and Wootton, J. C. (1990). Construction of validated, non-redundant composite protein sequence databases. *Protein Eng.* 3, 153–159. doi: 10.1093/protein/3.3.153

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780. doi: 10.1093/nar/27.8.1767

Bronstein, O., and Kroh, A. (2018). The first mitochondrial genome of the model echinoid *Lytechinus variegatus* and insights into Odontophoran phylogenetics. *Genomics* 111, 710–718. doi: 10.1016/j.ygeno.2018.04.008

Bullerwell, C. E., and Lang, B. F. (2005). Fungal evolution: the case of the vanishing mitochondrion. *Curr. Opin. Microbiol.* 8, 362–369. doi: 10.1016/j.mib.2005.06.009

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716. doi: 10.1016/j.tig.2003.10.012

Caspermeyer, J. (2016). MEGA evolutionary software re-engineered to handle today's big data demands. *Mol. Biol. Evol.* 33:1887.

Cheng, J., Luo, Q., Ren, Y. H., Luo, Z., Liao, W. L., Wang, X., et al. (2021). Panorama of intron dynamics and gene rearrangements in the phylum *Basidiomycota* as revealed by the complete mitochondrial genome of *Turbinellus floccosus*. *Appl. Microbiol. Biotechnol.* 105, 2017–2032. doi: 10.1007/s00253-021-11153-w

Coordinators, N. R. (2017). Database resources of the National Center for biotechnology information. *Nucleic Acids Res.* 44, D7–D19.

Dai, L. S., Kausar, S., Abbas, M. N., and Wang, T. T. (2018). Complete sequence and characterization of the *Ectropis oblique* mitochondrial genome and its phylogenetic implications. *Int. J. Biol. Macromol.* 107, 1142–1150. doi: 10.1016/j.ijbiomac.2017.09.093

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Dayo-Owoyemi, I., Rodrigues, A., Landell, M. F., Valente, P., Mueller, U. G., Ramos, J. P., et al. (2013). Intraspecific variation and emendation of *Hannaella kunmingensis*. *Mycol. Prog.* 12, 157–165. doi: 10.1007/s11557-012-0846-6

de Azeredo, L. A., Gomes, E. A., Mendonca-Hagler, L. C., and Hagler, A. N. (1998). Yeast communities associated with sugarcane in Campos, Rio de Janeiro, Brazil. *Int. Microbiol.* 1, 205–208.

Doyle, J. M., Bell, D. A., Bloom, P. H., Emmons, G., Fesnock, A., Katzner, T. E., et al. (2018). New insights into the phylogenetics and population structure of the prairie falcon (*Falco mexicanus*). *BMC Genomics* 19:233. doi: 10.1186/s12864-018-4615-z

du Toit, Z., du Plessis, M., Dalton, D. L., Jansen, R., Paul Grobler, J., and Kotze, A. (2017). Mitochondrial genomes of African pangolins and insights into evolutionary patterns and phylogeny of the family Manidae. *BMC Genomics* 18:746. doi: 10.1186/s12864-017-4140-5

Ferandon, C., Moukha, S., Callac, P., Benedetto, J. P., Castroviejo, M., and Barroso, G. (2010). The *Agaricus bisporus* cox1 gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *PLoS One* 5:e14048. doi: 10.1371/journal.pone.0014048

Ferandon, C., Xu, J., and Barroso, G. (2013). The 135 kbp mitochondrial genome of *Agaricus bisporus* is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. *Fungal Genet. Biol.* 55, 85–91. doi: 10.1016/j.fgb.2013.01.009

Fourie, G., Van der Merwe, N. A., Wingfield, B. D., Bogale, M., Wingfield, M. J., and Steenkamp, E. T. (2018). Mitochondrial introgression and interspecies recombination in the Fusarium fujikuroi species complex. *IMA Fungus* 9, 37–48. doi: 10.5598/imafungus.2018.09.01.04

Gan, H. M., Thomas, B. N., Cavanaugh, N. T., Morales, G. H., Mayers, A. N., Savka, M. A., et al. (2017). Whole genome sequencing of *Rhodotorula mucilaginosa* isolated from the chewing stick (*Distemonanthus benthamianus*): insights into *Rhodotorula phylogeny*, mitogenome dynamics and carotenoid biosynthesis. *PeerJ* 5:e4030. doi: 10.7717/peerj.4030

Glushakova, A. M., Iurkov, A. M., and Chernov, I. (2007). [Massive isolation of anamorphous ascomycete yeasts *Candida oleophila* from plant phyllosphere]. *Mikrobiologiia* 76, 896–901.

Gomes, F., Arantes, T. D., Fernandes, J. A. L., Ferreira, L. C., Romero, H., Bosco, S. M. G., et al. (2018). Polymorphism in mitochondrial group I introns among *Cryptococcus neoformans* and *Cryptococcus gattii* genotypes and its association with drug susceptibility. *Front. Microbiol.* 9:86. doi: 10.3389/fmicb.2018.00086

Hamari, Z., Juhasz, A., Gacser, A., Kucsera, J., Pfeiffer, I., and Kevei, F. (2001). Intron mobility results in rearrangement in mitochondrial DNAs of heterokaryon incompatible *Aspergillus japonicus* strains after protoplast fusion. *Fungal Genet. Biol.* 33, 83–95. doi: 10.1006/fgbi.2001.1272

Han, L., Li, Z. Y., Guo, X. F., Tan, J. L., He, S. Z., Cui, X. L., et al. (2017). *Hannaella dianchiensis* sp. nov., a basidiomycetous yeast species isolated from lake water. *Int. J. Syst. Evol. Microbiol.* 67, 2014–2018. doi: 10.1099/ijsem.0.001908

Kaewwichian, R., Jindamorakot, S., Am-In, S., Sipiczki, M., and Limtong, S. (2015). *Hannaella siamensis* sp. nov. and *Hannaella phetchabunensis* sp. nov., two new anamorphic basidiomycetous yeast species isolated from plants. *Int. J. Syst. Evol. Microbiol.* 65, 1297–1303. doi: 10.1099/ijs.0.000101

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

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 k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Kurtzman, C. P., Fell, F. W., Boekhout, T., and Robert, V. (2011). "Methods for isolation, phenotypic characterization and maintenance of yeasts," in *The Yeasts, a Taxonomic Study*, eds C. P. Kurtzman, J. W. Fell, and T. Boekhout (Amsterdam: Elsevier), 87–110. doi: 10.1016/b978-0-444-52149-1.00007-0

Landell, M. F., Billodre, R., Ramos, J. P., Leoncini, O., Vainstein, M. H., and Valente, P. (2010). *Candida aechmeae* sp. nov. and *Candida vrieseae* sp. nov., novel yeast species isolated from the phylloplane of bromeliads in Southern Brazil. *Int. J. Syst. Evol. Microbiol.* 60, 244–248. doi: 10.1099/ijs.0.011577-0

Landell, M. F., Brandao, L. R., Barbosa, A. C., Ramos, J. P., Safar, S. V., Gomes, F. C., et al. (2014). *Hannaella pagnoccae* sp. nov., a tremellaceous yeast species isolated from plants and soil. *Int J Syst Evol Microbiol* 64, 1970–1977. doi: 10.1099/ijs.0.059345-0

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773.

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397. doi: 10.1146/annurev.genet.33.1.351

Li, Q., Chen, C., Xiong, C., Jin, X., Chen, Z., and Huang, W. (2018a). Comparative mitogenomics reveals large-scale gene rearrangements in the mitochondrial genome of two *Pleurotus* species. *Appl. Microbiol. Biotechnol.* 102, 6143–6153. doi: 10.1007/s00253-018-9082-6

Li, Q., He, X., Ren, Y., Xiong, C., Jin, X., Peng, L., et al. (2020a). Comparative mitogenome analysis reveals mitochondrial genome differentiation in Ectomycorrhizal and Asymbiotic *Amanita* species. *Front. Microbiol.* 11:1382. doi: 10.3389/fmicb.2020.01382

Li, Q., Liao, M., Yang, M., Xiong, C., Jin, X., Chen, Z., et al. (2018b). Characterization of the mitochondrial genomes of three species in the ectomycorrhizal genus *Cantharellus* and phylogeny of Agaricomycetes. *Int. J. Biol. Macromol.* 118, 756–769. doi: 10.1016/j.ijbiomac.2018.06.129

Li, Q., Ren, Y., Shi, X., Peng, L., Zhao, J., Song, Y., et al. (2019a). Comparative mitochondrial genome analysis of two ectomycorrhizal fungi (Rhizopogon) reveals dynamic changes of intron and phylogenetic relationships of the Subphylum Agaricomycotina. *Int. J. Mol. Sci.* 20:5167. doi: 10.3390/ijms20205167

Li, Q., Ren, Y., Xiang, D., Shi, X., Zhao, J., Peng, L., et al. (2020b). Comparative mitogenome analysis of two ectomycorrhizal fungi (*Paxillus*) reveals gene rearrangement, intron dynamics, and phylogeny of basidiomycetes. *IMA Fungus* 11:12.

Li, Q., Wang, Q., Chen, C., Jin, X., Chen, Z., Xiong, C., et al. (2018c). Characterization and comparative mitogenomic analysis of six newly sequenced mitochondrial genomes from ectomycorrhizal fungi (Russula) and phylogenetic analysis of the Agaricomycetes. *Int. J. Biol. Macromol.* 119, 792–802. doi: 10.1016/j.ijbiomac.2018.07.197

Li, Q., Wang, Q., Jin, X., Chen, Z., Xiong, C., Li, P., et al. (2019b). Characterization and comparative analysis of six complete mitochondrial genomes from ectomycorrhizal fungi of the *Lactarius* genus and phylogenetic analysis of the Agaricomycetes. *Int. J. Biol. Macromol.* 121, 249–260. doi: 10.1016/j.ijbiomac.2018.10.029

Li, Q., Wu, P., Li, L., Feng, H., Tu, W., Bao, Z., et al. (2021). The first eleven mitochondrial genomes from the ectomycorrhizal fungal genus (Boletus) reveal intron loss and gene rearrangement. *Int. J. Biol. Macromol.* 172, 560–572. doi: 10.1016/j.ijbiomac.2021.01.087

Li, Q., Xiang, D., Wan, Y., Wu, Q., Wu, X., Ma, C., et al. (2019c). The complete mitochondrial genomes of five important medicinal *Ganoderma* species: features, evolution, and phylogeny. *Int. J. Biol. Macromol.* 139, 397–408. doi: 10.1016/j.ijbiomac.2019.08.003

Li, Q., Yang, L., Xiang, D., Wan, Y., Wu, Q., Huang, W., et al. (2020c). The complete mitochondrial genomes of two model ectomycorrhizal fungi (Laccaria): features, intron dynamics and phylogenetic implications. *Int. J. Biol. Macromol.* 145, 974–984. doi: 10.1016/j.ijbiomac.2019.09.188

Li, Q., Yang, M., Chen, C., Xiong, C., Jin, X., Pu, Z., et al. (2018d). Characterization and phylogenetic analysis of the complete mitochondrial genome of the medicinal fungus *Laetiporus sulphureus*. *Sci. Rep.* 8:9104.

Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41, W575–W581.

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57.

Molnarova, J., Vadkertiova, R., and Stratilova, E. (2014). Extracellular enzymatic activities and physiological profiles of yeasts colonizing fruit trees. *J. Basic Microbiol.* 54(Suppl. 1), S74–S84.

Perseke, M., Fritzsch, G., Ramsch, K., Bernt, M., Merkle, D., Middendorf, M., et al. (2008). Evolution of mitochondrial gene orders in echinoderms. *Mol. Phylogenet. Evol.* 47, 855–864. doi: 10.1016/j.ympev.2007.11.034

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* 34, 3299–3302. doi: 10.1093/molbev/msx248

Scheede-Bergdahl, C., and Bergdahl, A. (2017). Adaptation of mitochondrial expression and ATP production in dedifferentiating vascular smooth muscle cells. *Can. J. Physiol. Pharmacol.* 95, 1473–1479. doi: 10.1139/cjpp-2017-0227

Stajich, J. E., Wilke, S. K., Ahren, D., Au, C. H., Birren, B. W., Borodovsky, M., et al. (2010). Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc. Natl. Acad. Sci. U.S.A.* 107, 11889–11894.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Vaidya, G., Lohman, D. L., and Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27, 171–180. doi: 10.1111/j.1096-0031.2010.00329.x

Valach, M., Burger, G., Gray, M. W., and Lang, B. F. (2014). Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. *Nucleic Acids Res.* 42, 13764–13777. doi: 10.1093/nar/gku1266

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, J., Zhang, L., Zhang, Q. L., Zhou, M. Q., Wang, X. T., Yang, X. Z., et al. (2017). Comparative mitogenomic analysis of mirid bugs (Hemiptera: Miridae) and evaluation of potential DNA barcoding markers. *PeerJ* 5, e3661. doi: 10.7717/peerj.3661

Wang, Q. M., and Bai, F. Y. (2008). Molecular phylogeny of basidiomycetous yeasts in the *Cryptococcus luteolus* lineage (Tremellales) based on nuclear rRNA and mitochondrial cytochrome b gene sequence analyses: proposal of *Derxomyces* gen. nov. and *Hannaella* gen. nov., and description of eight novel *Derxomyces* species. *FEMS Yeast Res.* 8, 799–814. doi: 10.1111/j.1567-1364.2008.00403.x

Wang, X., Jia, L., Wang, M., Yang, H., Chen, M., Li, X., et al. (2020a). The complete mitochondrial genome of medicinal fungus *Taiwanofungus camphoratus* reveals gene rearrangements and intron dynamics of Polyporales. *Sci Rep* 10:16500.

Wang, X., Song, A., Wang, F., Chen, M., Li, X., Li, Q., et al. (2020b). The 206 kbp mitochondrial genome of *Phanerochaete carnosa* reveals dynamics of introns, accumulation of repeat sequences and plasmid-derived genes. *Int. J. Biol. Macromol.* 162, 209–219. doi: 10.1016/j.ijbiomac.2020.06.142

Wang, X., Wang, Y. J., Yao, W., Shen, J. W., Chen, M. Y., Gao, M., et al. (2020c). The 256 kb mitochondrial genome of *Clavaria fumosa* is the largest among phylum Basidiomycota and is rich in introns and intronic ORFs. *IMA Fungus* 11:26.

Wang, Z., Wang, Z., Shi, X., Wu, Q., Tao, Y., Guo, H., et al. (2018). Complete mitochondrial genome of Parasesarma affine (Brachyura: Sesarmidae): Gene rearrangements in Sesarmidae and phylogenetic analysis of the Brachyura. *Int. J. Biol. Macromol.* 118, 31–40. doi: 10.1016/j.ijbiomac.2018.06.056

Wu, P., Bao, Z., Tu, W., Li, L., Xiong, C., Jin, X., et al. (2021). The mitogenomes of two saprophytic *Boletales* species (Coniophora) reveals intron dynamics and accumulation of plasmid-derived and non-conserved genes. *Comput. Struct. Biotechnol. J.* 19, 401–414. doi: 10.1016/j.csbj.2020.12.041

Yadav, V., Sun, S., Billmyre, R. B., Thimmappa, B. C., Shea, T., Lintner, R., et al. (2018). RNAi is a critical determinant of centromere evolution in closely related fungi. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3108–3113. doi: 10.1073/pnas.1713725115

Yan, Z., Li, Z., Yan, L., Yu, Y., Cheng, Y., Chen, J., et al. (2018). Deletion of the sex-determining gene SXI1alpha enhances the spread of mitochondrial introns in *Cryptococcus neoformans*. *Mob DNA* 9, 24.

Yang, H., Xia, J., Zhang, J. E., Yang, J., Zhao, H., Wang, Q., et al. (2018). Characterization of the complete mitochondrial genome sequences of three croakers (Perciformes, Sciaenidae) and novel insights into the phylogenetics. *Int. J. Mol. Sci.* 19:1741. doi: 10.3390/ijms19061741

Yang, R. Y., Li, H. T., Zhu, H., Zhou, G. P., Wang, M., and Wang, L. (2012). Draft genome sequence of CBS 2479, the standard type strain of *Trichosporon asahii*. *Eukaryot. Cell* 11, 1415–1416. doi: 10.1128/ec.00237-12

Ye, J., Cheng, J., Ren, Y., Liao, W., and Li, Q. (2020). The first mitochondrial genome for Geastrales (*Sphaerobolus stellatus*) reveals intron dynamics and large-scale gene rearrangements of basidiomycota. *Front. Microbiol.* 11:1970. doi: 10.3389/fmicb.2020.01970

Yoon, H., Kong, W. S., Kim, Y. J., and Kim, J. G. (2016). Complete mitochondrial genome of the ectomycorrhizal fungus *Tricholoma matsutake*. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27, 3855–3857. doi: 10.3109/19401736.2014.958699

Zhang, S., Wang, X. N., Zhang, X. L., Liu, X. Z., and Zhang, Y. J. (2017). Complete mitochondrial genome of the endophytic fungus *Pestalotiopsis fici*: features and evolution. *Appl. Microbiol. Biotechnol.* 101, 1593–1604. doi: 10.1007/s00253-017-8112-0

Zhang, S., and Zhang, Y. J. (2019). Proposal of a new nomenclature for introns in protein-coding genes in fungal mitogenomes. *IMA Fungus* 10:15.

Zhang, Y., Sun, J., Rouse, G. W., Wiklund, H., Pleijel, F., Watanabe, H. K., et al. (2018). Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). *Mol. Phylogenet. Evol.* 125, 220–231. doi: 10.1016/j.ympev.2018.04.002

Zhang, Y. J., Zhang, H. Y., Liu, X. Z., and Zhang, S. (2017). Mitochondrial genome of the nematode endoparasitic fungus *Hirsutella vermicola* reveals a high level of synteny in the family Ophiocordycipitaceae. *Appl. Microbiol. Biotechnol.* 101, 3295–3304. doi: 10.1007/s00253-017-8257-x

Zhang, Y. J., Zhang, S., and Liu, X. Z. (2016). The complete mitochondrial genome of the nematode endoparasitic fungus *Hirsutella minnesotensis*. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27, 2693–2694. doi: 10.3109/19401736.2015.1046126

Zheng, B. Y., Cao, L. J., Tang, P., van Achterberg, K., Hoffmann, A. A., Chen, H. Y., et al. (2018). Gene arrangement and sequence of mitochondrial genomes yield insights into the phylogeny and evolution of bees and sphecid wasps (Hymenoptera: Apoidea). *Mol. Phylogenet. Evol.* 124, 1–9. doi: 10.1016/j.ympev.2018.02.028

Conflict of Interest: The 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 July 2021
doi: 10.3389/fmicb.2021.699598

Evidence for Persistent Heteroplasmy and Ancient Recombination in the Mitochondrial Genomes of the Edible Yellow Chanterelles From Southwestern China and Europe

Ying Zhang[1], Shaojuan Wang[1,2,3], Haixia Li[1,2], Chunli Liu[1,4], Fei Mi[1,5], Ruirui Wang[1,2], Meizi Mo[1,2] and Jianping Xu[1,6]**

[1] State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Key Laboratory for Southwest Microbial Diversity of the Ministry of Education, Yunnan University, Kunming, China, [2] School of Life Sciences, Yunnan University, Kunming, China, [3] Qicai Yunnan Primary School Affiliated with Yunnan Normal University, Kunming, China, [4] Kunming Edible Fungi Institute of All-China Federation of Supply and Marketing Cooperatives, Kunming, China, [5] Research Institute of Nutrition and Food Science, Kunming Medical University, Kunming, China, [6] Department of Biology, McMaster University, Hamilton, ON, Canada

OPEN ACCESS

Edited by:
Vassili N. Kouvelis,
National and Kapodistrian University of Athens, Greece

Reviewed by:
Akiyoshi Yamada,
Shinshu University, Japan
Dimitrios Floudas,
Lund University, Sweden

***Correspondence:**
Jianping Xu
jpxu@mcmaster.ca
Ying Zhang
yzh_1210@hotmail.com

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 23 April 2021
Accepted: 23 June 2021
Published: 14 July 2021

Citation:
Zhang Y, Wang S, Li H, Liu C, Mi F, Wang R, Mo M and Xu J (2021) Evidence for Persistent Heteroplasmy and Ancient Recombination in the Mitochondrial Genomes of the Edible Yellow Chanterelles From Southwestern China and Europe. Front. Microbiol. 12:699598. doi: 10.3389/fmicb.2021.699598

Mitochondrial genes and genomes have patterns of inheritance that are distinctly different from those of nuclear genes and genomes. In nature, the mitochondrial genomes in eukaryotes are generally considered non-recombining and homoplasmic. If heteroplasmy and recombination exist, they are typically very limited in both space and time. Here we show that mitochondrial heteroplasmy and recombination may not be limited to a specific population nor exit only transiently in the basidiomycete *Cantharellus cibarius* and related species. These edible yellow chanterelles are an ecologically very important group of fungi and among the most prominent wild edible mushrooms in the Northern Hemisphere. At present, very little is known about the genetics and population biology of these fungia cross large geographical distances. Our study here analyzed a total of 363 specimens of edible yellow chanterelles from 24 geographic locations in Yunnan in southwestern China and six geographic locations in five countries in Europe. For each mushroom sample, we obtained the DNA sequences at two genes, one in the nuclear genome and one in the mitochondrial genome. Our analyses of the nuclear gene, translation elongation factor 1-alpha (*tef-1*) and the DNA barcode of *C. cibarius* and related species, suggested these samples belong to four known species and five potential new species. Interestingly, analyses of the mitochondrial ATP synthase subunit 6 (*atp6*) gene fragment revealed evidence of heteroplasmy in two geographic samples in Yunnan and recombination within the two new putative species in Yunnan. Specifically, all four possible haplotypes at two polymorphic nucleotide sites within the mitochondrial *atp6* gene were found distributed across several geographic locations in Yunnan. Furthermore, these four haplotypes were broadly distributed across multiple

phylogenetic clades constructed based on nuclear *tef-1* sequences. Our results suggest that heteroplasmy and mitochondrial recombination might have happened repeatedly during the evolution of the yellow chanterelles. Together, our results suggest that the edible yellow chanterelles represent an excellent system from which to study the evolution of mitochondrial-nuclear genome relationships.

Keywords: DNA barcoding, speciation, biogeography, heteroplasmy, nuclear-mitochondrial incongruence, yellow chanterelles

INTRODUCTION

Fungi are important components of natural ecosystems. Many fungi also play significant roles in human and animal health, agriculture, biotechnology, and forestry. Some of these fungi grow in a mutually beneficial relationship with the root tips of plants, forming mycorrhizal associations. In these mycorrhizal fungi, their mycelia help plants obtain essential minerals, nitrogen and water from the soil and contribute to the plants' nutrition, disease resistance and drought tolerance (Brundrett, 2004). In return, plants provide their fungal partners with carbohydrates for their growth and reproduction. In forest ecosystems, one of the most common fungal-plant root associations (called mycorrhiza) involves the Basidiomycetes. Many of these basidiomycete fungi also produce conspicuous fruiting bodies—mushrooms. Some of these mushrooms, such as the chanterelles, are harvested as a source of highly prized food for humans.

Mitochondria are the powerhouse of all eukaryotes, responsible for generating the universal cellular energy currency ATP through oxidative phosphorylation. In fungi, previous studies have shown that mitochondria play many other roles, including aging in eukaryotes (Stojkovic et al., 2017), and virulence (Verma et al., 2018) and resistance to antifungal drugs in fungal pathogens (Vielba-Fernandez et al., 2018). In addition, mitochondrial genomes provide excellent materials for population and evolutionary studies. This is especially true in fungi where diverse patterns of mitochondrial inheritance have been observed, from uniparental to biparental inheritance (Zhou et al., 2010; de la Providencia et al., 2013; Xu and Li, 2015; Fourie et al., 2018; Mendoza et al., 2020). However, aside from a few model organisms, our knowledge about the mechanisms of fungal mitochondrial genomes and their inheritance is very limited (Basse, 2010; Wilson and Xu, 2012; Xu and Wang, 2015). In general, in contrast to the Mendelian inheritance of nuclear genes and genomes, the mitochondrial genes and genomes are inherited uniparentally with a preference for the maternal parent, with little or no evidence of recombination in most fungi, animals, and plants (Birky, 1995). As a consequence, all copies of the mitochondrial DNA (mtDNA) of an individual are typically identical to each other and to one of the parental mitochondrial genomes, a condition known as homoplasmy (Ling et al., 2011). Homoplasmy could be achieved through processes that prevent mtDNA transmission from one of the two parents, including no contribution from one parent due to fertilization mechanisms, targeted elimination of mtDNA from one parent through mitophagy, and/or the restriction-modification system (Shibata and Ling, 2007; Sato and Sato, 2011).

In filamentous fungi, the cells at the junctions of mating where the two parental homokaryons meet may be heteroplasmic, with each mated cell containing mtDNA from both parents [e.g., in the button mushroom *Agaricus bisporus* (Xu et al., 1996)]. In Baker's yeast *Saccharomyces cerevisiae*, the zygotes are almost universally heteroplasmic. However, in the human yeast pathogen *Cryptococcus neoformans*, heteroplasmy is only found in the zygotes when sex-determining genes are knocked out (Yan and Xu, 2003; Sun et al., 2020). In most instances of observed heteroplasmy, the heteroplasmic cells are typically transient, with rapid segregation of parental mtDNA genotypes into homoplasmic progeny cells to form homoplasmic mycelia and fruiting bodies (Birky, 2001; Barr et al., 2005; Wilson and Xu, 2012; Xu and Wang, 2015). However, while most reported heteroplasmy in fungi were from laboratory crosses, heteroplasmy has also been found in natural strains in a few basidiomycete fungi. For example, in our ongoing investigation of population genetics on edible basidiomycete mushrooms in southwestern China, over 87% of *Thelephora ganbajun* samples showed evidence of heterozygosity in the mitochondrial genes *cox1* or *cox3*, representing the first evidence of stable heteroplasmy in natural populations of basidiomycete fungi (Wang et al., 2017). In addition, we found clear evidences of phylogenetic incompatibility between two mitochondrial genes (*atp6* and *cox3*) among geographic populations of another basidiomycete mushroom, an *Russula virescens* ally (Cao et al., 2013). At present, the dynamics of heteroplasmy and recombination, while critical for advancing our knowledge of mitochondrial evolution, has not been critically investigated (Li et al., 2018).

The basidiomycete *Cantharellus cibarius* and related species are broadly found in the northern hemisphere, from Europe to North America and Asia (Buyck et al., 2014). These mushrooms can form symbiotic ectomycorrhizal relationships with many plants and are often found in mossy coniferous and birch forests. They are among the most prominent wild edible mushrooms in the northern hemisphere. Recently, phylogenetic analyses based on the translation elongation factor 1 (*tef-1*) has identified that the following 21 species *C. altipes*, *C. tenuithrix*, *C. lilacinopruinatus*, *C. ferruginascens*, *C. amethysteus*, *C. lewisii*, *C. lateritius*, *C. confluens*, *C. spectaculus*, *C. roseocanus*, *C. flavus*, *C. phasmatis*, *C. isabellinus*, *C. tomentosus*, *C. avellaneus*, *C. tabernensis*, *C. appalachiensis*, *C. decolorans*, *C. texensis*, *C. cinnabarinus*, were phylogenetically related to *C. cibarius*

(Buyck and Hofstetter, 2011; Foltz et al., 2013; Buyck et al., 2014). However, most of these species appear to be geographically specific. For example, *C. cibarius* sensu stricto is only found in Scandinavia and northern Japan (Ogawa et al., 2018). Therefore, the idea that Europe and the warm-temperate and subtropical parts of North America have species of *Cantharellus* in common, has been progressively abandoned. Consequently, it's increasingly recognized that the use of European species names of *Cantharellus* for those found in North America and Asia is likely incorrect, especially for those closely related to *C. amethysteus* and *C. cibarius* (Redhead et al., 1998; Buyck and Hofstetter, 2011). At present, while updates on *Cantharellus* taxonomy have progressed relatively quickly in western Europe and North America, that in Asia remains very fragmented. However, recent studies have reported several new chanterelles from China (Shao et al., 2011; Tian et al., 2012; Shao et al., 2014, 2016a,b; An et al., 2017; Jian et al., 2020), Japan (Suhara and Kurogi, 2015; Ogawa et al., 2018), Korea (Antonin et al., 2017; Buyck et al., 2020), Malaysia (Eyssartier et al., 2009; Buyck et al., 2014), India (Das et al., 2015; Buyck et al., 2018), and Iran (Parad et al., 2018). Recently, a large-scale survey of wild edible mushrooms in local markets in Yunnan province, southwestern China, revealed that there was at least one cryptic species within this group of species in this region (Zhang et al., 2021). This study analyzed 96 samples using sequence information at both the *tef-1* locus and the internal transcribed spacer (ITS) regions of the ribosomal RNA gene cluster. However, a very conservative cutoff for species delimitation was used in new species estimation in this study (Zhang et al., 2021). Another study identified that most of the edible yellow chanterelles old in the Yunnan local markets belonged to *C. Yunnanensis* based on both morphological and *tef-1*sequences, and not the traditionally assumed *C. cibarius* (Shao et al., 2021). Unfortunately, the rRNA gene cluster often fail to distinguish closely related species in *Cantharellus* due to unreadable sequence chromatographs and/or limited DNA sequence variations. Sequence information from other genes as well as more extensive sampling is needed to obtain a better understanding of the evolutionary history of *Cantharellus* and better insights on dispersal routes and speciation events in the Northern Hemisphere (Olariaga et al., 2017).

Chanterelle is the common name for mushrooms in the genus *Cantharellus* and in other morphologically related genera. The edible yellow chanterelles are among the most popular edible wild mushrooms in the world and include *C. cibarius* and related species. There are specialized companies that commercialize fruit bodies of chanterelles of different sizes and presentations, e.g., as fresh, frozen, dry, or salted, with geographic specifications. Morphologically, species in yellow chanterelles are very similar to each other and all species are harvested as food and some of them are globally traded. The combined global commercial value of yellow chanterelles has been estimated at more than one billion US dollars annually (Watling, 1997; Wang and Hall, 2004). One of the main global centers of production and harvesting of edible yellow chanterelles is Yunnan Province in Southwestern China. Due to its highly variable climate and diverse topography, Southwestern China (mainly Yunnan Province) is recognized as one of the world's 34 biodiversity hotspots (Myers et al., 2000). Relevant to this study, Yunnan is also one of the world's most important areas for wild mushroom harvesting and trading. For example, of the 2,000 or so wild edible mushroom species in the world, over 900 are found in Yunnan (Yang, 2002; Feng and Yang, 2018). Based on ITS barcoding, our recent survey identified a high species diversity of wild mushrooms in local markets, including a large number of putative new taxa (Zhang et al., 2021). Some of these mushrooms are very important to both the local and regional economy. Most of these species are ectomycorrhizal, contribute significantly to the health of trees and forest ecosystems. Examples of these economically important ectomycorrhizal mushrooms around the globe include *Boletus edulis, Thelephora ganbajun, Tricholoma matsutake, Russula virescens, Russulavinosa* and *C. cibarius* etc.

In this study, we analyzed the DNA sequence variation in the mitochondrial *atp6* gene fragment for a total of 363 mushroom samples of edible yellow chanterelles. Our analyses found that two nucleotide sites within this gene fragment contained sequence heterogeneity within two individual fruiting bodies and that all four possible genotypes at these two sites were found in several geographic populations from Yunnan. Similarly, three additional heterozygous loci were found in another fruiting body collected in a different geographic population. Interestingly, seven out of all eight possible haplotypes at these three nucleotide sites were found in our samples, consistent with recombination. Finally, the phylogenetic distributions of two polymorphic nucleotide sites (sites 101 and 109) within the *atp6* gene fragment were compared with sequence variation at the DNA barcode *tef-1* for edible yellow chanterelles in Yunnan and Europe.

MATERIALS AND METHODS

Mushroom Sampling, DNA Isolation and Gene Sequencing

In this study, we obtained 219 fruiting bodies with macro-morphological characteristics similar to *C. cibarius* and related species from 24 counties in Yunnan province. The sampled areas in Yunnan spanned about 480 km from east to west and about 550 km from north to south. In addition, we obtained 144 yellow chanterelle samples from local markets in five countries (Finland, France, Hungary, Portugal, and Russia) in Europe. These 363 samples were all included in the study. The geographic coordinates and the sample size from each site are presented in **Table 1**.

DNA extraction, PCR, sequencing, and sequence alignment all followed those of Wang et al. (2017). These samples were first analyzed based on their sequences at the *tef-1* gene, the preferred DNA barcode for *Cantharellus* (Buyck and Hofstetter, 2011; Antonin et al., 2017; Olariaga et al., 2017). In our study, DNA sequences at the nuclear*tef-1*gene were used to confirm phylogenetic affiliations of all our specimens (see below). In addition, the mitochondrial *atp6* gene sequences were obtained for 327 specimens. The primer pairs, tef-1-S (5′ ACCAGAACGACGCCGCTCAA 3′) and tef-1-A (5′ TTGTCGCCGTGCCAACCAG 3′), ATP6-S (5′ AACACCGGGTACATTTCT 3′) and ATP6-A

TABLE 1 | Distribution and diversity of *tef-1* sequence types for yellow chanterelles across southwestern China and Europe.

Geographic population (province)/ collecting (city)	Latitude (°N)	Longitude (°W)	sample size	ST(no. of isolates in each ST)
Russia (1)	55.45	37.37	27	12(1) 13(2) 14(2) 16(22)
France (2)	45.28	4.2	15	6(1) 14(4) 16(10)
Portugal (3, 4)	38.3	28	65	14(8) 16(45) 17(3) 37(1) 38(8)
Hungary (5)	47.26	19.15	12	2(1) 8(1) 16(10)
Southern France (6)	41.18	2.06	10	1(1) 3(1) 8(1) 16(7)
Finland (8, 9)	60.13	24.5	15	10(1) 11(1) 14(2) 15(1) 16(10)
Changning (CN)	24.67	102.14	3	4(1) 45(2)
Dali (DL)	25.69	100.19	9	5(8) 45(1)
Fengqing (FQ)	24.58	99.91	5	45(5)
Jiangcheng (JC)	22.58	101.88	12	9(1) 19(1) 27(1) 31(2) 41(1) 45(6)
Jianchuan (JChuan)	25.82	100.55	16	25(1) 28(1) 30(1) 31(5) 32(1) 40(1) 45(6)
Jinning (JN)	23.88	102.58	9	23(1) 25(1) 31(3) 35(2) 45(2)
Kunming (KM)	25.04	102.73	5	22(1) 28(1) 33(1) 34(1) 36(1)
Lancang (LAN)	23.38	100.55	1	5(1)
Lijiang (LJ)	26.88	100.25	8	7(1) 35(3) 39(1) 45(3)
Lincang (LC)	25.21	100.09	9	40(3) 45(6)
Lufeng (LF)	25.15	101.26	1	5(1)
Menghai (MH)	21.96	100.45	20	5(3) 31(1) 40(5) 45(11)
Nanhua (NH)	25.55	101.26	8	23(1) 31(1) 35(5) 47(1)
Nanjian (NJ)	25.04	100.51	1	31(1)
Pu'er (PE)	23.33	100.50	1	29(1)
Shangyun (SY)	22.35	99.98	5	5(5)
Shizong (SZ)	24.83	103.98	21	26(1) 27(6) 28(7) 31(2) 32(1) 35(3) 45(1)
Shuangjiang (SJ)	23.45	99.85	7	5(3) 44(1) 45(3)
Xundian (XD)	24.67	103.25	19	18(2) 21(1) 31(1) 40(1) 45(13) 46(1)
Xiangyun (XY)	25.47	100.56	12	26(1) 31(2) 35(7) 45(1) 46(1)
Yimen (YM)	24.57	101.00	23	18(3) 20(1) 40(2) 43(1) 45(14) 46(1)
Yunxian (YX)	24.44	100.12	2	4(1) 24(1)
Yongping (YP)	24.67	102.14	2	5(2)
Zhongdian (ZD, A)	27.78	100.97	20	7(15) 22(1) 24(1) 35(3)
Total			363	

(5′ TACTTACGGCGATTTCTA 3′) were used to amplify and sequences the *tef-1* and *atp6* gene fragments, respectively.

mtDNA Heterozygosity

We observed mtDNA heterozygosity, shown as double peaks at specific nucleotide sites in the chromatograms of the *atp6*gene, in several specimens. We investigated whether the observed mtDNA heterozygosity was due to nuclear mt paralogs (numts) or the presence of more than one allele in the mitogenome within each of these specimens. To distinguish these two possibilities, we used two approaches. In the first, we confirmed the nucleotide sequence similarity between our sequences and that of the sequenced genome of *C. cibarius*/sample MG75 (GenBank number: QOWL00000000.1) through Basic Local Alignment Search Tool (BLAST). [Note: Based on its *tef-1* sequence, sample MG75 likely belongs to *C. tuberculosporus*.] The second approach was through copy number assessment using a technique called absolute quantification PCR (AQ-PCR) (Leong et al., 2007).

In the second approach, for samples with evidence of heterogeneity at the *atp6* gene, we determined their relative abundances in the fruiting bodies between the two combinations, one combination with peaks (TT) and the second combination with peaks (CA) respectively (**Figure 1A**) using AQ-PCR. Their abundances were compared with that of a nuclear gene, β-tubulin (*tub*). Since chanterelle mushrooms are diploids, each cell contains two haploid nuclei and multiple mitochondrial genomes. Thus, there would be two copies of each nuclear gene such as *tub* within each cell while multiple copies of mitochondrial genes. In this case, the relative copy numbers of the two *atp6* alleles can be compared with the copy number of *tub* to determine within each sample whether the *atp6* alleles are located in the nuclear genome or the mitochondrial genome. Our specific protocols are described as follows.

To detect the relative copy numbers of the two *atp6* alleles, we first amplified the *atp6* gene fragment using primers ATP6H1 (5′ TGTAGCTAGAGCTTTCTCTTTAGGAGT 3′) and ATP6H2 (5′CCACTTGTAAATAATTTAGTTAAAAGACC3′). Similarly, the nuclear reference gene β-tubulin was amplified using primers TUB1 (5′CAGGAGGGTATGGATGAG3′) and TUB2 (5′TAACTGGAGGGGAGAATG3′). The amplified products containing the DNA targets were excised from 1% agarose gel under UV illumination and purified, then confirmed by sequencing to ensure the unique polymorphism is included in each *atp6* products. Furthermore, the concentration of the amplified products was measured with a spectrophotometer, using the average molecular weight of the product and Avogadro's constant. The copy number per unit volume was then calculated by comparing to a standard curve.

To establish the standard curve for gene copy number in a sample, both the *tub* and *atp6* genes were first cloned separately using the PMD18-T cloning vector kit (CAT. 6011, TaKaRa, JAPAN) according to the manufacturer's instructions. The recombinant vector was transformed into competent *Escherichia coli* cells and 25 μL of transformed culture was spread onto LB plates containing ampicillin (75 μg/ml) and X-gal/IPTG (CAT. R1171, ThermoFisher, United States). Transformed (white) colonies were picked and processed for plasmid isolation. Plasmid purification was done using a Plasmid Mini kit (CAT. TP01100, GENERAL BIOSYSTEMS, China). The presence of the insert in the recombinant clones was confirmed by restriction digestion. The cloned circular plasmid was quantified using a spectrophotometer and linearized with restriction enzyme *Hind* III (CAT.ER0501, ThermoFisher, United States). The stock solutions of linearized plasmid DNAs were serially diluted to obtain a standard series from 10^9 to 10 copies per μL with each step differing by 10 folds. Then, standard curves were

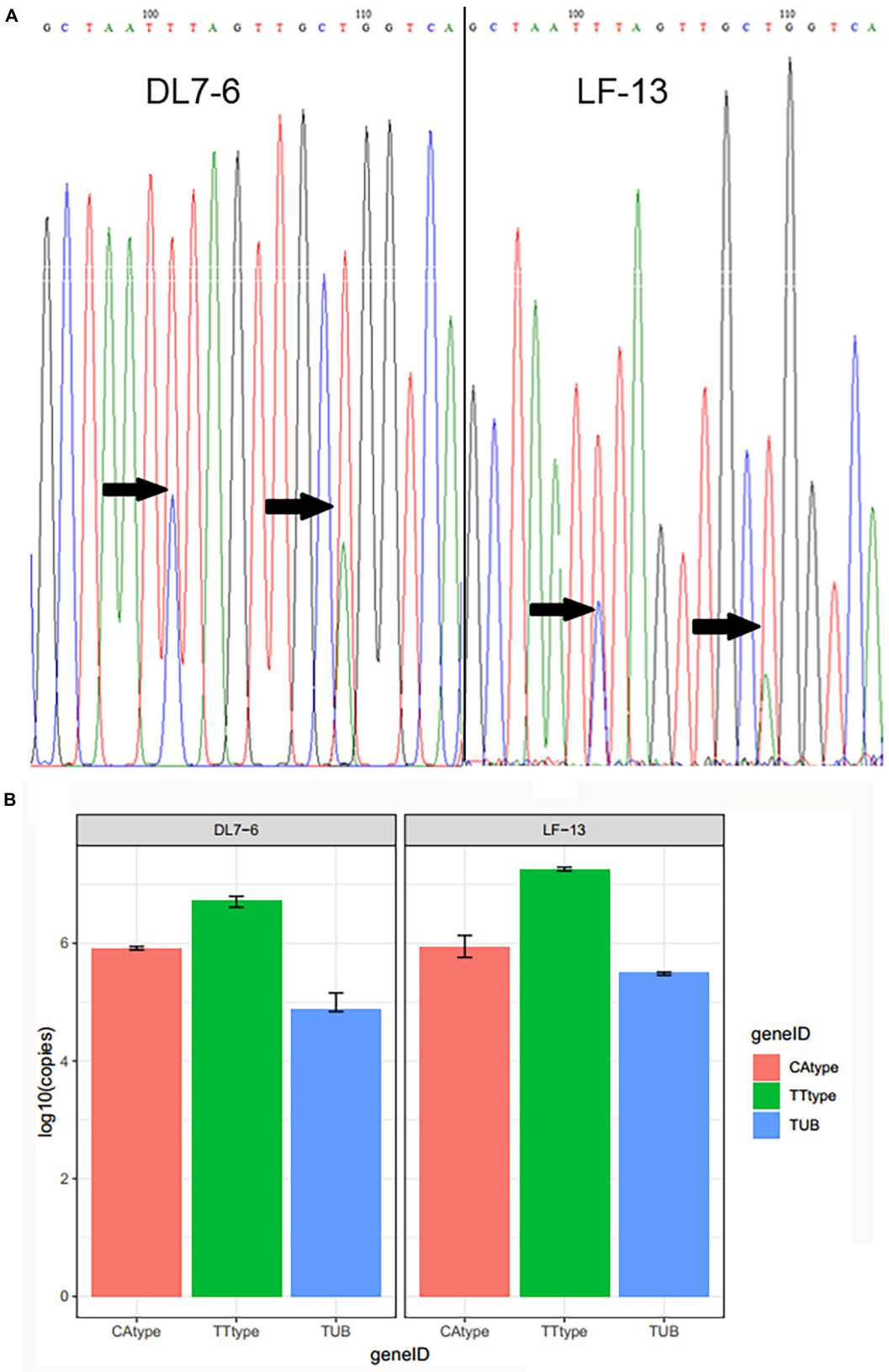


FIGURE 1 | (A) Heterozygous sites of mitochondrial gene *atp6* and **(B)** copy numbers of higher peak and lower peak sequences relative to the reference nuclear *tub* gene in samples DL7-6 and LF-13 from Yunnan.

built using 6 serial dilutions of each of the three plasmid DNA containing three different DNA fragments, *atp6*-TT, *atp6*-CA and *tub*. For AQ-PCR, the corresponding standards series was run under the same conditions and the copy numbers of each gene in each specimen was calculated using the following formula (Godornes et al., 2007):

$$\frac{6.022 \times 10^{23}\left(\frac{\text{molecules}}{\text{mole}}\right) \times \text{DNA concentrations}\left(\frac{\text{g}}{\mu\text{l}}\right)}{\text{Number of bases pairs} \times 660 \text{ daltons}}$$

6.022×10^{23} (molecules/mole) Avogadro's number
660 Da: the average weight of a single base pair.

After the standard curves are established, we used quantitative-PCR (qPCR) to determines the actual copy numbers of target genes by relating the Ct values of each gene in each specimen to the standard curves (Xue et al., 2014). Each qPCR reaction was performed three times using the CFX96® system (Bio-Rad, United States) that contained the AceQ qPCR TaqMan Master Mix (without ROX) (CAT.Q112, Vazyme, China), 10 uM of dual labeled hydrolysis probe (Probe-TT: FAM-TTTAGTTGCTGGTCACACATTACT-MGB for the TT combination at nucleotide sites 101 and 109 of *atp6*, probe-CA: FAM-TCTAGTTGCAGGTCATACATTATT-MGB for the CA combination at nucleotide sites 101 and 109 of *atp6*, and probe-TUB:FAM-AGCACATACACGGCATCCTGG-BHQ1 at *tub* locus), and 10 uM each of forward and reverse primers. During each round of PCR, the target and reference sequences were simultaneously amplified by AmpliTaq® Gold DNA Polymerase. This enzyme has a 5′ nuclease activity that cleaves probes that are hybridized to each amplicon sequence. When an oligonucleotide probe is cleaved by the AmpliTaq Gold DNA Polymerase 5′ nuclease activity, the quencher is separated from the reporter dye, and thus increasing the fluorescence of the reporter dye. Accumulation of PCR products can be detected in real time by monitoring the increase in fluorescence of each reporter dye at each PCR cycle. Ct, slope, PCR efficiency, correlation coefficient (R2) and percentage of variance in copy numbers were calculated by using the default settings of Bio-Rad CFX Manager Version 3.1 (Bio-Rad, United States).

Data Analysis

Species Identification

Because of the relatively high heterogeneity of the primary fungal DNA barcode ITS within individual samples of *C. cibarius* and related species, high quality ITS sequences are often not available for species identification within this species complex (Buyck et al., 2020; Zhang et al., 2021). Instead, as in previous studies (Buyck et al., 2011; Buyck and Hofstetter, 2011), the *tef-1* gene was used as the DNA barcode to determine the taxonomic placements of our *Cantharellus* specimens. However, at present, a *tef-1* sequence-based new species identification system is not yet available. Because the systematics and taxonomy of Asian *Cantharellus* remains largely unresolved, for this study, we chose *tef-1* sequence variations among samples within each of several closely related known species to establish a cutoff value for defining putative phylogenetic species limits.

To establish the cutoff value, we first retrieved all known *tef-1* sequences of *C. cibarius* and related species from GenBank. The following 13 known species within *C. cibarius* and related species were found to contain *tef-1* sequences from three or more specimens each: *C. anzutake*, *C. tabernensis, C. appalachiensis, C. tenuithrix,C. ferruginascens, C. cinnabarinus, C. cibarius s.s., C. tuberculosporus, C.texensis, C.decolorans, C. amethysteus, C. lateritius,* and *C. lewisii*. The distributions of these eight species have been well-documented and their morphological features and phylogenetic placements are known (Buyck et al., 2014). These sequences were used to identify the range of intraspecific *tef-1* sequence variation and establish a conservative sequence-based cutoff value for putative phylogenetic species identification within the yellow chanterelles. Here, the highest intraspecific variation within these eight species was chosen as our cutoff value to define species limits. Subsequently, *tef-1* sequence variations between our samples and those published previously for the known species of the yellow chanterelles were compared to the cutoff value. Specifically, we consider any specimen with a greater divergence than the cutoff from all known species and from each other in our sample as a potential new phylogenetic species. In contrast, specimens with *tef-1* sequence divergence lower than the cutoff value were considered as belonging to the same species.

For each of our specimens, DNA sequences obtained from both forward and reverse primers of the *tef-1* gene were assembled using the SeqMan sequence analysis software (DNASTAR, Inc.). All *tef-1* sequences obtained for our specimens and those from GenBank representing the diversity of species within the yellow chanterelles were aligned by using MAFFT 6.0 (Misawa et al., 2002) and checked manually by BioEdit 7.0.9 (Hall, 1999). Ambiguous positions at the two ends of each gene fragment were excluded from the analysis. For the *tef-1* dataset, maximum likelihood (ML) and Bayesian inference (BI) analyses were performed using RAxML 7.2.6 (Stamatakis, 2006) and MrBAYES3.1.2 (Huelsenbeck and Ronquist, 2001). The best partition schemes and evolutionary models were selected using MrModeltest v2.3. ML analyses were run with all parameters set to the default settings. GTR+G was used as the most appropriate model and the bootstrap analysis run with 1,000 replicates. BI analysis consisted of four simultaneous Markov chain Monte Carlo (MCMC) chains was run over 5×10^6 generations with trees sampled every 100 generations when the average standard deviation of split frequencies was lower than 0.01. By omitting the first 25% of trees as burn-ins using the "sump" and "sumt" commands, a majority rule consensus tree was generated. The percentage of sequence divergence for each specimen from all other specimens and known species were calculated and used to define the taxonomic placements of each of our specimens.

Haplotype Inference of *tef-1* and *atp6* Genotype Distribution

In this study, like most other mushrooms, we consider each individual mushroom as derived from a mating event between two genetically different homokaryons. Thus, each mushroom is a heterokaryon, with each cell having two different haploid nuclei and each locus having two alleles. In the case of single nucleotide polymorphisms (SNPs), each nucleotide site

in each individual specimen may be homozygous for a specific nucleotide or heterozygous, containing two different nucleotides. In the sequenced *tef-1* fragment, there are multiple SNP sites within many specimens. To better infer the relationships among specimens, we also inferred the *tef-1* haplotypes for each specimen using the Bayesian method implemented in the program PHASE 2.1 (Stephens and Donnelly, 2003). The inferred haplotype sequences were then imported into MrBAYES 3.2 to analyze the relationships among alleles from within the same and different mushroom fruiting bodies, the*atp6* sequence of each sample was assigned to one of the five *atp6* genotypes at the two polymorphic nucleotide sites (positions 101 and 109): heterozygous genotype with both CA and TT or homozygous genotypes CA, TA, CT, or TT. For specimens containing haplotypes with ambiguous phylogenetic placements, we also cloned their *tef-1* alleles and sequenced each cloned allele separately, following the protocol described for obtaining pure sequences of *atp6*and *tub*. The generated phylogenetic relationships were analyzed together with their geographic origins to identify geographic patterns of (putative) species distributions.

RESULTS

In this study, we analyzed a total of 363 mushroom specimens of the yellow chanterelles from 30 geographic locations in six countries. 24 of the locations were in Yunnan province in China while the remaining six were in Europe. The six European geographic locations had at least 10 specimens in each, with a range of 10 to 65. Among the 24 geographic locations from Yunnan, the sample sizes ranged from 1 to 23; with four having only one specimen each, 12 having 2 to 9 specimens each, and eight having 10 or more specimens each (**Table 1**). Each of these specimens were analyzed for its nuclear *tef-1* and mitochondrial *atp6* sequences. Below we describe the patterns of sequence variation among these specimens, with an emphasis on mitochondrial DNA polymorphisms.

tef-1 Genotyping of the 363 Yellow Chanterelle Specimens

Our sequence analyses identified that out of the 303 aligned nucleotide sites among the 363 *tef-1* sequences, 123 sites were polymorphic. These 123 SNPs resolved the 363 specimens into 47 *tef-1* sequence types (STs) (Peakall and Smouse, 2012). Among the 47 STs, 24 were each shared by more than one specimen each while the remaining 23 were found only once each. The most frequent ST was sequence type16, shared by 104 specimens from seven geographic populations. The second most frequent ST was sequence type45, shared by 75 specimens from 15 geographic populations. Among the 30 geographic samples, 23 contained more than one ST each while the remaining seven had only one ST in each local population. However, four of the seven geographic samples with only one ST each had only one specimen each, with the remaining three geographic samples each having two to five specimens. Samples from two geographic locations from Jianchuan and Shizhong contained most *tef-1* STs, both including seven STs each (**Table 1**).

Phylogenetic Reconstruction and Phylogenetic Species Recognition

For phylogenetic analyses, we retrieved the *tef-1* sequences of *C. cibarius* and those closely related species as revealed based on previous studies (Buyck and Hofstetter, 2011; Foltz et al., 2013; Buyck et al., 2014). Specifically, except *C. confluens* which didn't have any *tef-1*sequence deposited in GenBank, all known species of chanterelles were included in our comparisons. In addition, *C. anzutake, C. tuberculosporus,* and *C. yunnanensis,* which were recently described as very closely related species to *C. cibarius* s.s. and some of our Yunnan samples (Ogawa et al., 2018; Shao et al., 2021), was also included in the comparisons. In total,115 *tef-1* sequences were included in our phylogenetic analyses, with 47 sequences representing the 47 unique STs found in our current samples while the remaining 68 sequences were retrieved from NCBI to represent the diverse species within or closely related to *C. cibarius* s.s.

The ML and BI analyses of *tef-1* dataset yielded identical tree topologies for the 115 *tef-1* sequences. The tree inferred from BI analysis is shown in **Figure 2**. As shown in **Supplementary Table 1**, among the 21 known reference species, *C. anzutake* showed the highest value (0.8%) of intra-specific variation of *tef-1* sequences. Thus, 0.8% sequence variation at *tef-1* locus was chosen as the cutoff value for our putative phylogenetic species identification using *tef-1* sequences within this group of chanterelles. Taking this value as a threshold, the largest interspecific genetic distances between *C. tenuithrix*, *C. phasmatis*, and *C. flavus* (0.1%) were all lower than 0.8%. Furthermore, these three species formed a single clade in our *tef-1* phylogeny. Two of these species, *C.phasmatis* and *C. flavus*, were shown to share a more recent common ancestor than with the recently described species from the southern United States *C. tenuithrix* (Foltz, 2011). Therefore, based on the 0.8% *tef-1* sequence divergence criteria to define putative phylogenetic species, *C. tenuithrix*, *C. phasmatis,* and *C. flavus* would be combined as one species/species complex. Similarly, genetic distance between *C. ferruginascens* and *C. lilacinopruinatus* was 0.1%. Indeed, both *C. ferruginaseens* and *C. lilacinopruinatus* are characterized by having a white stipe with similar pileus margin and hymenophore (Olariaga and Salcedo, 2008). Thus, these two species could be combined as one species too. The same applies to *C. roseocanus* and *C. cibarius*, which were recently separated from each other as distinct species (Foltz, 2011; Thorn et al., 2017). Thus, our criterion of requiring greater than 0.8% sequence divergence at the *tef-1* locus represents a conservative cutoff value that minimizes the number of potential new species in our collection. Based on this conservative criterion, there are likely nine putative species in our samples. Among these nine putative species, two (*Cantharellus* sp.1 and *Cantharellus* sp. 4) clustered with known species *C. tuberculosporus* and *C. cibarius,* respectively. Two others, *Cantharellus* sp. 6 and *Cantharellus* sp. 8, have <0.8% of genetic distance with the phylogenetically related species

C. amethysteus and *C. lewisii*, respectively. The remaining five had genetic divergences from known species at a much higher than the 0.8% cutoff value, thus likely representing new putative phylogenetic species (**Figure 2**, **Supplementary Tables 1**, **2**). Among the proposed new putative species, two (*Cantharellus* sp. 2 and 5) were only based on one sample each, while the other three were each based on more than one samples. Thus, to formally establish these putative species as new species, more samples need to be collected and analyzed, especially for sp. 2 and 5, in their natural habitats. Regardless of the final number of species, these results suggest a significant number of undescribed species and genotypic diversity in our analyzed yellow chanterelles.

Evidence for Heteroplasmy Within Mitochondrial *atp6*

We successfully obtained *atp6* sequences from 327 of the 363 specimens. When we examined the *atp6* sequence chromatographs, two specimens, DL7-6 and LF-13 (both belonging to *Cantharellus* sp. 9), were found to have double peaks at sites 101 and 109, while another specimen SY6-4 had double peaks at sites 62, 95, and 132, consistent with heterozygosity at these sites. At all sites, the lower peaks were significantly higher than the baseline noise levels at neighboring homozygous sites within the chromatograph (**Figure 1A**). To exclude the possibility of cross-contamination, we re-sampled DNA from additional tissues of these three specimens and amplified and sequenced the *atp6* gene fragment again to confirm. For controls, genomic DNA from two other samples (ZD12-13 and YM3-30, both are in *Cantharellus* sp. 1) that initially did not show any double peaks in their *atp6* chromatographs were also re-extracted and re-sequenced. Our sequence analyses confirmed the double peaks at the same nucleotide sites in the additional tissues of DL7-6, LF-13, and SY6-4, but no evidence of heterozygosity was found in specimens ZD12-13 and YM3-30. Our results excluded the possibility of sample contamination as a cause for the observed double peaks in specimens DL7-6, LF-13, and SY6-4. In both the original chromatographs and the new chromatographs, the remaining 301 nucleotides of the sequenced *atp6* gene fragment each had only one nucleotide in specimens DL7-6, LF-13, and SY6-4, with no evidence of heterozygosity. Below, we systematically analyzed the distribution and variation patterns of heterozygosity at these sites (101 and 109). The three heterozygous sites found in specimen SY6-4 were separated from sites 101 and 109 in the following phylogenetic analyses of *atp6* gene sequences.

Comparisons using BLAST showed that both nucleotide sequences with the higher and lower peaks were most similar to the *atp6* gene from *Cantharellus cibarius* sample MG75 (GenBank number: QOWL00000000.1). The sequence identity of the *atp6* haplotype with the higher peaks to *atp6* gene of sample MG75 was 94% (E-value is 5e-099), and that with the lower peaks was 95% (E-value is 1e-101). The results are consistent with both sequences (one has the combination of the high peaks TT and the other has the combination of low peaks CA) are homologous sequences from the mitochondrial genome.

Our copy number analyses confirmed that both the TT allele and the CA allele of the *atp6* gene are more numerous than the nuclear *tub* gene in both specimens DL7-6 and LF-13. **Figure 1B** summarized the copy numbers of the CA and TT haplotypes and of the nuclear *tub* gene from specimens DL7-6 and LF-13. Specifically, in specimen DL7-6, the CA and TT allele copy numbers were 10.9- and 69.9-fold higher than *tub*. Similarly, in specimen LF-13, the CA and TT allele copy numbers were significantly higher than that of *tub* gene, at 2.7- and 57.1- folds of *tub*. The differential ratios of the two alleles of *atp6* gene both within each specimen and between the two specimens are also consistent with their non-nuclear nature (Sandor et al., 2018).

Distribution of atp6 Genotypes Along the *tef-1* Phylogeny

The phylogenetic distributions of *atp6* haplotypes at the two nucleotide sites (sites 101 and 109) were mapped onto the Bayesian tree constructed based on the *tef-1* dataset (**Figure 3**). The results showed that the four different haplotype combinations (CA, CT, TA, and TT) at these two sites were broadly distributed throughout the *tef-1* sequence-based phylogenetic tree. For example, among the European samples of the yellow chanterelles, three haplotypes CA, CT, and TT were found at these two nucleotide sites. Similarly, in multiple clades of specimens from Yunnan, three to four haplotypes out of all four possible haplotypes were found at these two *atp6* nucleotide sites. Specifically, except for two clades *Cantharellus* sp. 2 and *Cantharellus* sp. 5 (both with just one sample each, **Figure 2**), other seven putative phylogenetic species recognized by *tef-1* sequences have two or more different *atp6* haplotype combinations. Within *Cantharellus* sp. 1 and *Cantharellus* sp. 9, all four haplotypes at the two sites were found, consistent with recombination within this mitochondrial gene fragment in the population. However, if true, during subsequent evolution and diversification, multiple parallel mutations and/or reversions at the two sites must have occurred to account for the observed patterns of distribution. However, given the low sequence diversity among the samples at *atp6* gene fragment, we believe that the chance of parallel mutation or reversal at exactly these two sites in multiple clades is extremely small. The second possibility is that mitochondrial recombination is frequent within and among contemporary populations of the yellow chanterelles throughout their distribution range, a single recombination event from the heteroplasmic parent followed by segregation could have generated all four possible genotype combinations at these two sites. However, this second possibility requires most of the proposed species to hybridize with each other to generate recombinant mitochondrial genotypes. The third possibility is that the ancestor of the yellow chanterelles was heteroplasmic and recombined to produce four haplotypes between these two polymorphic nucleotide sites. These recombinants and heteroplasmic individuals then diversified among descendent lineages but with incomplete

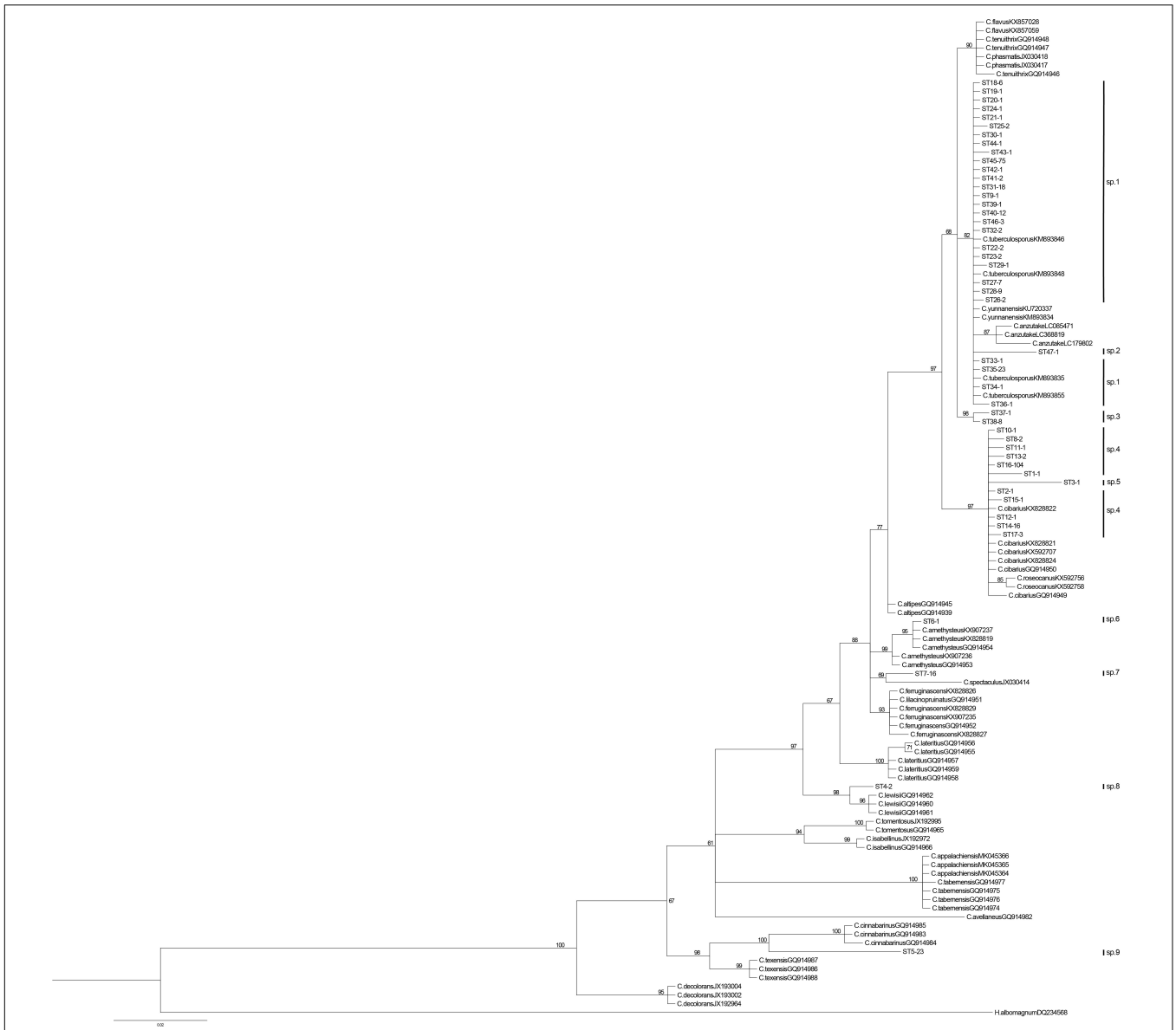


FIGURE 2 | *tef-1* phylogenetic tree of the *C. cibarius* species complex. For each *tef-1* sequence type (ST) in our samples, the name or abbreviation (full name can be found in **Supplementary Table 2**) represents the county/community from where the sample was obtained, the number after represents the ST code assignment; the last number represents the total number of samples belonging to the specific ST.

lineage sorting, resulting in some lineages maintaining the heteroplasmic state while others containing one to four of the four possible haplotypes.

To further investigate evidence of recombination among polymorphic nucleotide sites within the mitochondrial *atp6* gene, we measured the degree of association between alleles at seven polymorphic nucleotide sites within this DNA fragment using the program Multilocus 1.3 (Agapow and Burt, 2001). The analyses were conducted for both the total samples and samples from individual phylogenetically distinct clusters as revealed by the *tef-1* phylogeny. Only clades with >5 samples were analyzed. Because I_A can be influenced by the number of polymorphic loci and samples, we standardized the I_A value by the rBarD value for comparisons among populations. The null hypothesis for I_A is that there is random association (recombination) among alleles at different loci, a p value of <0.05 would indicate that the null hypothesis should be rejected. Phylogenetic incompatibility (PrC) is another indicator of recombination at the population level, the lack of phylogenetic incompatibility implies no recombination (Agapow and Burt, 2001). Results of I_A and PrC tests showed that while random recombination was rejected in all samples, evidence for recombination was found in both the total sample as well as in *Cantharellus* sp. 9 and *C. tuberculosporus* (**Table 2**). Specifically, they showed

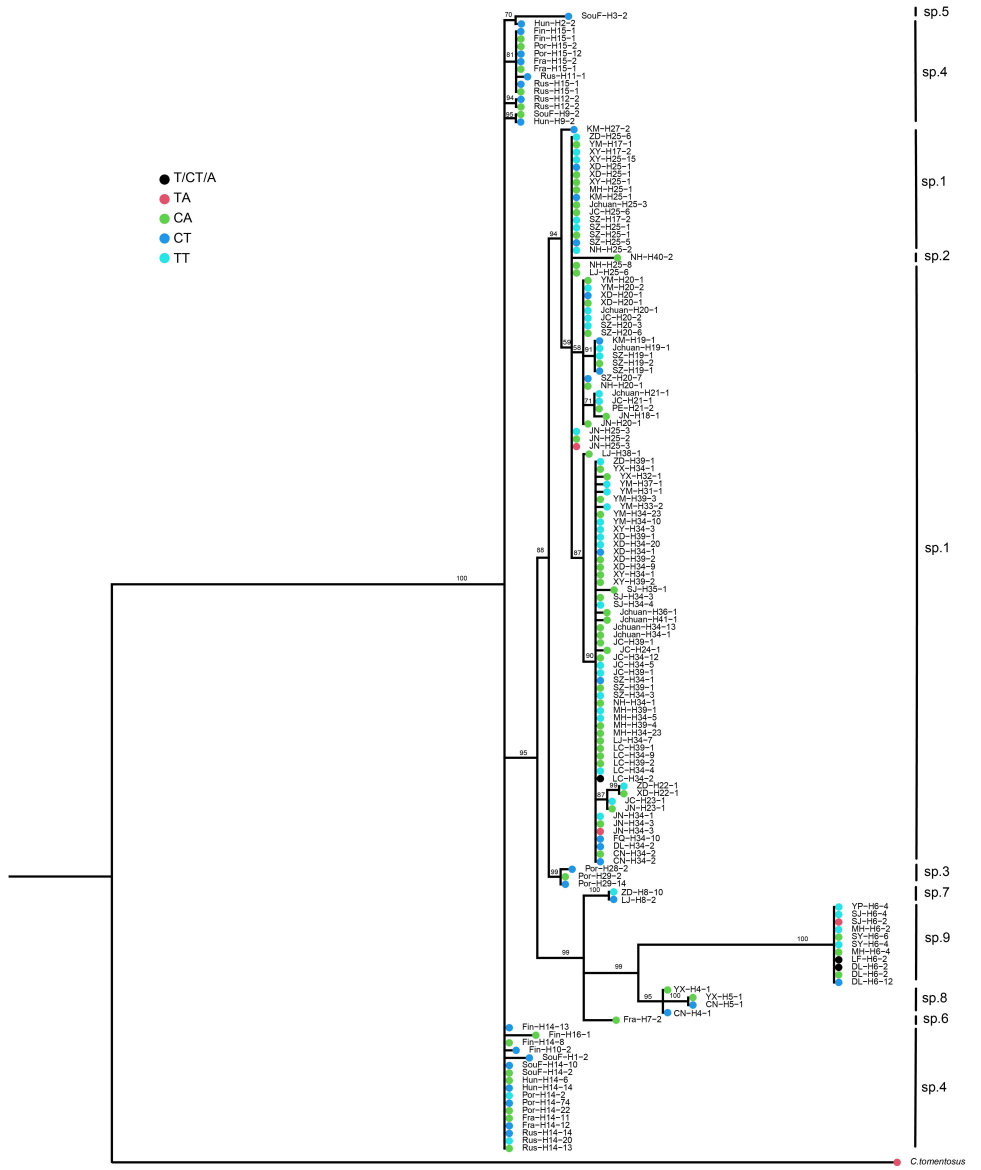


FIGURE 3 | Bayesian *tef-1* haplotype cladogram showing the broad distributions of*atp6* genotypes at two polymorphic nucleotide sites. Termina branch colors represent the different *atp6*genotypes at the two polymorphic sites; For each *tef-1* haplotype in our samples, the name or abbreviation (full name can be found in **Supplementary Table 2**) represents the county/community from where the sample was obtained, the number after "H" represent the haplotype assignment; the last number represents the total number of samples belonging to the specific haplotype from each geographic location and for each unique *atp6* genotype. Only representative sequences of unique haplotypes from each geographic location and unique *atp6* genotypes are shown.

clear evidence of phylogenetic incompatibility (**Table 2**). Overall, our results are consistent with non-random recombination in the mitochondrial genome of several distinct clades within the yellow chanterelles.

Using the same seven non-heteroplasmic sites within the *atp6* gene fragment, we constructed the phylogenetic relationships of all clone-corrected samples and compared it with that constructed using *tef-1* SNPs. The phylogenetic comparisons revealed evidence for phylogenetic in congruences between the two gene trees among most phylogenetic species. An example is strain SY6-4 that showed distinctly incongruent phylogenetic placements between the *tef-1* and *atp6* phylogenies. The results further support recombination and hybridization within and between several clades within the yellow chanterelles.

DISCUSSION

Species and Genetic Diversity

In this study, 363 fruiting bodies of the edible yellow chanterelles were obtained from Yunnan, Southwestern China and five countries in Europe. Phylogenetic analyses of our *tef-1* sequence

TABLE 2 | Association among SNPs at the *atp6* gene for *C. cibarius* and related species across southwestern China and Europe.

Species groups	Sample size	Phylogenetic compatibility (p value)	rBarD (p value)
All samples	327	0.7826 (< 0.001)	0.1827 (< 0.001)
Cantharellus sp. 3	9	1 (0.035)	1 (0.035)
C. tuberculosporus[1]	158	0.9710 (< 0.001)	0.3004 (< 0.001)
Cantharellus sp. 7	5	1 (1)	n/a[3]
Cantharellus sp. 9	14	0.8986 (< 0.001)	0.3488 (< 0.001)
C. cibarius[2]	132	1(1)	0.3726 (< 0.001)

1. Known species from Yunnan; 2. Known species from Europe; 3. Not analyzed due to small sample size.

types and closely related sequences from GenBank identified that while many of our samples belonged to two known species *C. cibarius* (most of our European samples) and *C. tuberculosporus* (many of our Yunnan samples), our specimens belonged to seven additional putative phylogenetic species.

Previous studies have shown that the broad application of the name *C. cibarius* to chanterelles found in North America and other areas outside of Europe were inaccurate. As a result, several new species have been described (Buyck and Hofstetter, 2011). The phylogenetic analyses of our samples from Southwest China along with representative samples of *C. cibarius* and related species from other parts of world indicated that the Yunnan and European samples contained a total of nine putative phylogenetic species. However, due to the lack of ITS, *tef-1,* and other gene sequences in public databases for several known species in this group of mushrooms, we are unable to confidently finalize the phylogenetic species status of the four clades. Indeed, though ITS shows higher resolution for species identification after cloning step (Ogawa et al., 2018), the difficulties in the direct amplification and sequencing of this gene, as well as the shortage of ITS data in the public database due to its high frequency of heterogeneity in many *Cantharellus* samples, have made ITS not an ideal barcode locus for this group of fungi (Xu, 2016). Instead, *tef-1* has been recognized as the DNA barcode for *C. cibarius* and related species (Olariaga et al., 2017).

By constructing the *tef-1* phylogenetic tree with the reference sequences representing the few current known species within the edible yellow chanterelles, we made several novel observations. First, several known species distinguished based on morphological characters showed no or very limited difference between each other at the *tef-1* gene fragment. Indeed, the lack of a clear barcode gap at the *tef-1* gene among several groups of morphological species suggest that additional genes or whole-genome sequences should be used to confirm the genetic uniqueness of these species (Xu, 2020). Second, most of the samples from Yunnan had distinct *tef-1* sequences from those from other parts of the world. For example, most of the samples from Yunnan belonged to one clade that clustered with *C. anzutake, C. tuberculosporus*, and *C. yunnanensis* (Ogawa et al., 2018; Shao et al., 2021; Zhang et al., 2021). However, until recently, the common chanterelles in Yunnan province were called *C. cibarius*, the species found only in Europe (Liu et al., 2009; Wang et al., 2009). The known species *C. tuberculosporus* and *C. yunnanensis* showed very low interspecies genetic distance (0.3%) from each other, and with only 0.8% between *C. yunnanensis* and *C. anzutake*. According to the cutoff value we established for species delimitation in this study, the three species could be combined into one monophyletic species. Thus, our analyses indicate that the name *C. cibarius* is a misnomer when applied based on morphological characteristics to chanterelles in Yunnan, a conclusion similar to what was found recently through morphological investigations (Shao et al., 2021). With such a stringent criterion, several specimens from Zhongdian (*Cantharellus* sp. 7) and Nanhua (*Cantharellus* sp. 2) in Yunnan most likely belonged to novel species. Further comparisons using morphological characteristics and additional DNA sequences are needed to confirm their genetic uniqueness and cryptic speciation.

Persistent Heteroplasmy and Ancient Mitochondrial Recombination

We found evidence for heteroplasmy and recombination within the mitochondrial gene *atp6* in our samples of *C. cibarius* and related species. The identification of heterozygous nucleotide sites at the *atp6* locus in samples DL7-6 and LF-13 is consistent with heteroplasmy. In laboratory crosses, heteroplasmy is a pre-requisite for mitochondrial recombination (Basse, 2010). Furthermore, according to the *tef-1* phylogeny, although the two heteroplasmic specimens DL7-6 and LF-13 belonged to the same *tef-1* clade, the four *atp6* haplotypes as observed from the yellow chanterelles had wide geographic and phylogenetic distributions, indicating that recombination was likely ancient. For recombination to occur, heteroplasmy must have been present in the ancestor to drive the recombination event. However, during subsequent evolution and diversification, most lineages may have lost heteroplasmy and/or some of the recombinant *atp6* genotypes. The exceptions are two specimens DL7-6 and LF-13 that maintained their heteroplasmy.

Advances in DNA sequencing and polymorphism detection technology have documented the presence of recombinant mtDNA for a wide range of eukaryotic taxa. In the Baker's yeast *Saccharomyces cerevisiae*, there is a large body of data showing evidence of heteroplasmy and recombination in laboratory crosses. However, mitochondrial heterogeneity was always transient, with the offspring of the zygote becoming completely homozygous within 20 mitotic cell divisions (Hermann and Shaw, 1998). A few other studies have also provided evidence of mtDNA recombination in natural populations of certain fungal species, such as *Armillaria gallica* (Saville et al., 1998), *Cryptococcus gattii* (Xu et al., 2009), *Agaricus bisporus* (Xu et al., 2013) and recently, *Thelephora ganbajun* (Wang et al., 2017). However, there has been no report showing evidence of potential stable heteroplasmy and broad mitochondrial recombination over repeated speciation events in nature.

In the human pathogenic yeast *C. neoformans*, a progeny population with diverse mitochondrial genotypes created due to biparental mitochondrial inheritance was found to show greater phenotypic variation than that with a uniparental mitochondrial

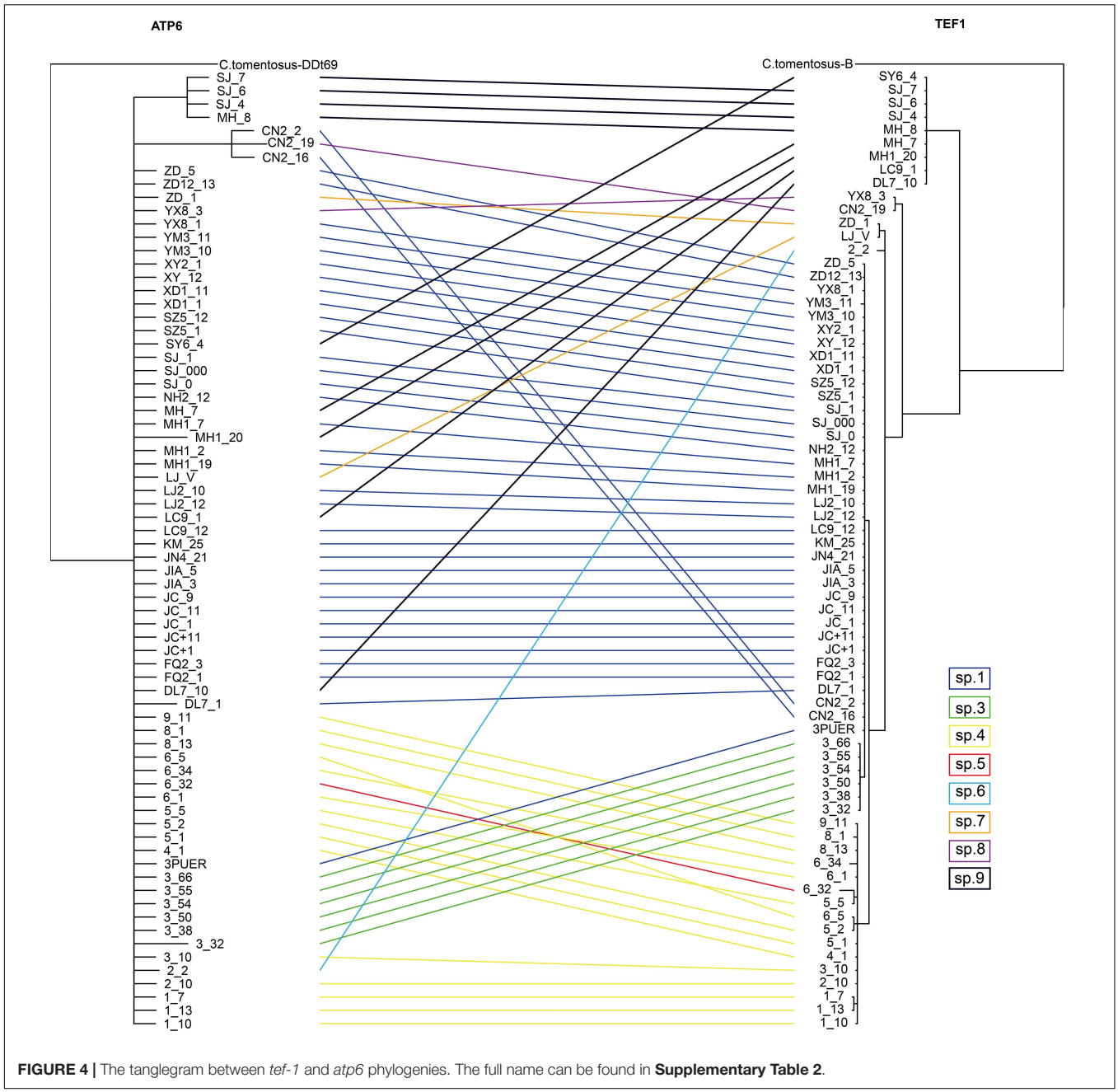


FIGURE 4 | The tanglegram between *tef-1* and *atp6* phylogenies. The full name can be found in **Supplementary Table 2**.

inheritance (Yan et al., 2007). Mitochondrial genomes are commonly inherited uniparentally in the majority of sexual eukaryotes. As a result, these genomes are considered effectively asexual that may be prone to Muller's Rachet, an irreversible accumulation of deleterious mutations (Lynch, 1996; Xu, 2005). Interestingly, our analyses identified signatures of heteroplasmy and mtDNA recombination in the yellow chanterelles, suggesting these organisms may have an effective mechanism to prevent Muller's Rachet from operating in their mitochondrial genomes (Matsumoto and Fukumasa-Nakai, 1996; Xu, 2007).

At present, the selective pressure responsible for the maintenance of heteroplasmy and mitochondrial recombination in the yellow chanterelles is not known. In the human pathogenic yeast *C. neoformans*, high temperatures and UV exposure have shown to be capable of changing mitochondrial inheritance from uniparental to biparental (Yan et al., 2007), which also included mitochondrial heteroplasmy and the generation of mitochondrial recombinants. The high altitudes in Yunnan where these mushroom samples are from may be exposed to significant UV irradiations or other stress factors that favor mitochondrial heteroplasmy and recombination as an adaptive response to reduce the rate of accumulation of deleterious mutations in the mitochondrial genomes in these species. Indeed, evidence for mitochondrial recombination have been found in

several other mushroom species (e.g., *T. ganbajun* and a *Russula virescens* ally) in the same geographic regions (Cao et al., 2013; Wang et al., 2017). Further investigations of the relationships between environmental factors and mitochondrial genetic variations are needed to determine the extent of mitochondrial recombination and the potential adaptive significance of such recombination in these mushrooms.

Comparison Between Mitochondrial and Nuclear Genetic Polymorphisms

In our analyses, there were overall lower sequence divergences among clades at the *atp6* locus than those at *tef-1* locus. This result is consistent with what has been reported in many other fungi that showed lower rates of evolution in the mitochondrial genome than that in the nuclear genome (Sandor et al., 2018, Sandor et al., 2020). Indeed, distinct genotypes and lineages identified based on the nuclear *tef-1* sequences were often found to share the same *atp6* sequences. However, the lower sequence diversity for mitochondrial genes than nuclear genes in these mushrooms doesn't mean that the mitochondrial genomes in these fungi are static. Contrary, the mitochondrial genome size variation among strains in the same or closely related species can be very large (up to $\pm$25% of the mean genome size), usually much greater than the nuclear genome size variation within many fungal species (Sandor et al., 2018; Wang and Xu, 2020; Zhang et al., 2020). The mitochondrial genome size variations in fungi were mainly due to differences in the number and size of introns and mobile genetic elements (Sandor et al., 2018; Wang and Xu, 2020; Zhang et al., 2020). Indeed, in our initial screening for mitochondrial genetic markers for analyzing the yellow chanterelles, polymorphisms in intron distributions were found in several genes, including *cox1, cox2, cox3, nad1, nad5, cob,* and *rnl* (**Supplementary Table 3** and unpublished data). Unfortunately, such intron distribution polymorphisms in these genes limited the success of our direct PCR and sequencing efforts for these specimens. Consequently, these gene fragments were not selected as markers to further investigate the mitochondrial genetic polymorphisms in our study.

We would like to mention that while the low rate of evolution at the mitochondrial *atp6* gene may be responsible for some of the haplotype sharing, an alternative explanation is recent hybridization among some of the lineages. Indeed, our analyses revealed evidence of incongruence between the nuclear *tef-1* phylogeny and the mitochondrial *atp6* phylogeny, consistent with hybridization among lineages. Analyses of whole-genome sequences, including both the nuclear and mitochondrial genome sequences at the population level, is needed in order to distinguish these two possibilities.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the GenBank database, accession number(s) can be found in **Supplementary Table 2**.

AUTHOR CONTRIBUTIONS

YZ and JX: Conceptualization, validation, visualization, supervision, project administration, and funding acquisition. JX, SW, and MM: sample collection. SW, HL, CL, FM, RW, and MM: methodology. SW and HL: software and formal analysis. YZ, JX, and SW: investigation. YZ and SW: writing—original draft preparation. JX: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was jointly supported by the National Natural Science Foundation of China (31870009 and 31760010) to YZ, the "Double First Class" Research Project in Yunnan University (C176280107) to JX, and the Top Young Talents Program of the Ten Thousand Talents Plan in Yunnan Province to YZ.

ACKNOWLEDGMENTS

We thank Vassili N. Kouvelis and Georg Hausner for the invitation. We thank the following people for their help in sample collections: Xiaozhao Tang, Yang Cao, Yanchun Li, Gang Wu, Lei Zhang, Xin Zhang, Junfeng Liang, Zaochang Liu, Shanze Yoell, and Liam Yoell.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.699598/full#supplementary-material

Supplementary Table 1 | Average evolutionary divergence over *tef-1* sequence pairs within and between groups (provincially adopted phylogenetic species) calculated by MEGA 6 using the K2P model.

Supplementary Table 2 | List of putative species, corresponding ST code and samples used in **Figures 2**, **4**, *tef-1* haplotype code, corresponding *atp6* genotype and sample origins used in **Figure 3**.

Supplementary Table 3 | Intron numbers of mitochondrial genes in *C. cibarius* and related species.

REFERENCES

Agapow, P. M., and Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Mol. Ecol. Resour.* 1, 101–102. doi: 10.1046/j.1471-8278.2000.00014.x

An, D. Y., Liang, Z. Q., Jiang, S., Su, M. S., and Zeng, N. K. (2017). Cantharellus hainanensis, a new species with a smooth hymenophore from tropical China. *Mycoscience* 58, 438–444. doi: 10.1016/j.myc.2017.06.004

Antonin, V., Hofstetter, V., Ryoo, R., Ka, K. H., and Buyck, B. (2017). New *Cantharellus* species from the Republic of Korea. *Mycol. Prog.* 16, 753–759.

Barr, C. M., Neiman, M., and Taylor, D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and

animals. *New Phytol.* 168, 39–50. doi: 10.1111/j.1469-8137.2005.01492.x

Basse, C. W. (2010). Mitochondrial inheritance in fungi. *Curr. Opin. Microbiol.* 13, 712–719. doi: 10.1016/j.mib.2010.09.003

Birky, C. W. Jr. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. U.S.A.* 92, 11331–11338. doi: 10.1073/pnas.92.25.11331

Birky, C. W. Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35, 125–148. doi: 10.1146/annurev.genet.35.102401.090231

Brundrett, M. (2004). Diversity and classification of mycorrhizal associations. *Biol. Rev.* 79, 473–495. doi: 10.1017/S1464793103006316

Buyck, B., Antonín, V., Chakraborty, D., Baghela, A., Das, K., and Hofstetter, V. (2018). Cantharellus sect. Amethystini in Asia. *Mycol. Prog.* 17, 917–924. doi: 10.1007/s11557-018-1403-8

Buyck, B., Cruaud, C., Couloux, A., and Hofstetter, V. (2011). *Cantharellus texensis* sp. nov. from Texas, a southern lookalike of *C. cinnabarinus* revealed by tef-1 sequence data. *Mycologia* 103, 1037–1046. doi: 10.3852/10-261

Buyck, B., and Hofstetter, V. (2011). The contribution of tef-1 sequences to species delimitation in the *Cantharellus cibarius* complex in the southeastern USA. *Fungal Divers.* 49, 35–46. doi: 10.1007/s13225-011-0095-z

Buyck, B., Kauff, F., Eyssartier, G., Couloux, A., and Hofstetter, V. (2014). A multilocus phylogeny for worldwide Cantharellus (Cantharellales, Agaricomycetidae). *Fungal Divers.* 64, 101–121. doi: 10.1007/s13225-013-0272-3

Buyck, B., Valerie, H., Ryoo, R., Ka, K.-H., and Antonín, V. (2020). New *Cantharellus species* from South Korea. *MycoKeys* 76, 31–47. doi: 10.3897/mycokeys.76.58179

Cao, Y., Zhang, Y., Yu, Z., Mi, F., Liu, C., Tang, X., et al. (2013). Structure, gene flow, and recombination among geographic populations of a *Russula virescens* Ally from southwestern China. *PLoS One* 8:e73174. doi: 10.1371/journal.pone.0073174

Das, K., Hofstetter, V., Chakraborty, D., Baghela, A., Singh, S. K., and Buyck, B. (2015). *Cantharellus sikkimensis* sp nov. (Cantharellales, Agaricomycetes) from the Indian Himalayas. *Phytotaxa* 222, 267–275. doi: 10.11646/phytotaxa.222.4.4

de la Providencia, I. E., Nadimi, M., Beaudet, D., Morales, G. R., and Hijri, M. (2013). Detection of a transient mitochondrial DNA heteroplasmy in the progeny of crossed genetically divergent isolates of arbuscular mycorrhizal fungi. *New Phytol.* 200, 211–221. doi: 10.1111/nph.12372

Eyssartier, G., Stubbe, D., Walleyn, R., and Verbeken, A. (2009). New records of *Cantharellus* species (Basidiomycota, Cantharellaceae) from Malaysian dipterocarp rainforest. *Fungal Divers.* 36, 57–67. doi: 10.1002/yea.1667

Feng, B., and Yang, Z. L. (2018). Studies on diversity of higher fungi in Yunnan, southwestern China: a review. *Plant Divers.* 40, 165–171. doi: 10.1016/j.pld.2018.07.001

Foltz, M. J. (2011). *Systematics and Molecular Phylogeny of Cantharellus spp. in Western Wisconsin*. Ph. D. thesis. La Crosse, WI: University of Wisconsin-La Crosse.

Foltz, M. J., Perez, K. E., and Volk, T. J. (2013). Molecular phylogeny and morphology reveal three new species of *Cantharellus* within 20 m of one another in western Wisconsin, USA. *Mycologia* 105, 447–461. doi: 10.3852/12-181

Fourie, G., Van Der Merwe, N. A., Wingfield, B. D., Bogale, M., Wingfield, M. J., and Steenkamp, E. T. (2018). Mitochondrial introgression and interspecies recombination in the *Fusarium fujikuroi* species complex. *IMA Fungus* 9, 37–48. doi: 10.5598/imafungus.2018.09.01.04

Godornes, C., Leader, B., Molini, B., Centurion-Lara, A., and Lukehart, S. (2007). Quantitation of rabbit cytokine mRNA by real-time RT-PCR. *Cytokine* 38, 1–7. doi: 10.1016/j.cyto.2007.04.002

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98. doi: 10.1021/bk-1999-0734.ch008

Hermann, G., and Shaw, J. (1998). Mitochondrial dynamics in Yeast. *Annu. Rev. Cell Dev. Bio.* 14, 265–303. doi: 10.1146/annurev.cellbio.14.1.265

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Jian, S., Dai, R., Gao, J., and Feng, B. (2020). *Cantharellus albus*, a striking new species from Southwest China. *Phytotaxa* 470, 133–144. doi: 10.11646/phytotaxa.470.2.2

Leong, D. T., Gupta, A., Bai, H. F., Wan, G., Yoong, L. F., Too, H. P., et al. (2007). Absolute quantification of gene expression in biomaterials research using real-time PCR. *Biomaterials* 28, 203–210. doi: 10.1016/j.biomaterials.2006.09.011

Li, Q., Liao, M., Yang, M., Xiong, C., Jin, X., Chen, Z., et al. (2018). Characterization of the mitochondrial genomes of three species in the ectomycorrhizal genus *Cantharellus* and phylogeny of Agaricomycetes. *Int. J. Biol. Macromol.* 118, 756–769. doi: 10.1016/j.ijbiomac.2018.06.129

Ling, F., Mikawa, T., and Shibata, T. (2011). Enlightenment of yeast mitochondrial homoplasmy: diversified roles of gene conversion. *Genes* 2, 169–190. doi: 10.3390/genes2010169

Liu, P., Wang, X., Yu, F., Chen, J., Tian, X., Deng, X., et al. (2009). Fungous kingdom: Yunnan of China and her Ectomycorrhizal macrofungal species diversity. *Acta Botanica Yunnanica Suppl* 31, 15–20.

Lynch, M. (1996). Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol. Biol. Evol.* 13, 209–220. doi: 10.1093/oxfordjournals.molbev.a025557

Matsumoto, T., and Fukumasa-Nakai, Y. (1996). Mitochondrial DNA inheritance in sexual crosses of *Pleurotus ostreatus*. *Curr. Genet.* 30, 549–552. doi: 10.1007/s002940050168

Mendoza, H., Perlin, M., and Schirawski, J. (2020). Mitochondrial inheritance in phytopathogenic fungi-everything is known, or is it? *Int. J. Mol. Sci.* 21:3883. doi: 10.3390/ijms21113883

Misawa, K., Katoh, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acid. Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi: 10.1038/35002501

Ogawa, W., Endo, N., Fukuda, M., and Yamada, A. (2018). Phylogenetic analyses of Japanese golden chanterelles and a new species description, *Cantharellus anzutake* sp nov. *Mycoscience* 59, 153–165. doi: 10.1016/j.myc.2017.08.014

Olariaga, I., Moreno, G., Manjon, J. L., Salcedo, I., Hofstetter, V., Rodriguez, D., et al. (2017). *Cantharellus* (Cantharellales, Basidiomycota) revisited in Europe through a multigene phylogeny. *Fungal Divers.* 83, 263–292. doi: 10.1007/s13225-016-0376-7

Olariaga, I., and Salcedo, I. (2008). *Cantharellus ilicis* sp. nov., a new species from the Mediterranean Basin collected in evergreen Quercus forests. *Rev. Catalana Micología* 30, 107–116.

Parad, G. A., Ghobad-Nejhad, M., Tabari, M., Yousefzadeh, H., Esmaeilzadeh, O., Tedersoo, L., et al. (2018). *Cantharellus alborufescens* and *C. ferruginascens* (Cantharellaceae, Basidiomycota) new to Iran. *Crypto. Mycol.* 39, 299–310. doi: 10.7872/crym/v39.iss3.2018.299

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. population genetic software for teaching and research–an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Redhead, S., Norvell, L. L., and Danell, E. (1998). *Cantharellus formosus* and the Pacific golden chanterelle harvest in western North America. *Mycotaxon (USA)* 65, 285–322. doi: 10.1017/S0953756297003997

Sandor, S., Zhang, Y.-J., and Xu, J. (2018). Fungal mitochondrial genomes and genetic polymorphisms. *Appl. Microb. Biotech.* 102, 1–16. doi: 10.1007/s00253-018-9350-5

Sandor, S. R., Wang, H. C., Vaario, L. M., Trudell, S. A., and Xu, J. (2020). Mitochondrial multilocus DNA sequence analyses reveal limited genetic variability within and consistent differences Between species within the global matsutake species complex. *Acta Edulis. Fungi.* 27, 1–19. doi: 10.16488/j.cnki.1005-9873.2020.01.001

Sato, M., and Sato, K. (2011). Degradation of paternal mitochondria by fertilization-triggered autophagy in *C. elegans* embryos. *Science* 334, 1141–1144. doi: 10.1126/science.1210333

Saville, B. J., Kohli, Y., and Anderson, J. B. (1998). mtDNA recombination in a natural population. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1331–1335. doi: 10.1073/pnas.95.3.1331

Shao, S. C., Buyck, B., Hofstettere, V., Tian, X. F., Geng, Y. H., Yu, F. Q., et al. (2014). Cantharellus hygrophorus, a new species in subgenus *Afrocantharellus*

from tropical southwestern China. *Crypto. Mycol.* 35, 283–291. doi: 10.7872/crym.v35.iss3.2014.283

Shao, S. C., Buyck, B., Tian, X. F., Liu, P. G., and Geng, Y. H. (2016a). *Cantharellus phloginus*, a new pink-colored species from southwestern China. *Mycoscience* 57, 144–149. doi: 10.1016/j.myc.2015.12.004

Shao, S.-C., Liu, P.-G., Tian, X., Buyck, B., and Geng, Y.-H. (2016b). A new species of *Cantharellus* (Cantharellales, Basidiomycota, Fungi) from subalpine forest in Yunnan, China. *Phytotaxa* 252:273. doi: 10.11646/phytotaxa.252.4.3

Shao, S.-C., Liu, P.-G., Wei, T.-Z., and Herrera, M. (2021). New insights into the taxonomy of the genus *Cantharellus* in China: epityfication of *C. yunnanensis* W.F. Chiu and the first record of *C. cibarius* Fr. *Cryptogam. Mycol.* 42, 25–37. doi: 10.5252/cryptogamie-mycologie2021v42a3

Shao, S. C., Tian, X. F., and Liu, P. G. (2011). *Cantharellus* in southwestern China: a new species and a new record. *Mycotaxon* 116, 437–446. doi: 10.5248/116.437

Shibata, T., and Ling, F. (2007). DNA recombination protein-dependent mechanism of homoplasmy and its proposed functions. *Mitochondrion* 7, 17–23. doi: 10.1016/j.mito.2006.11.024

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Stephens, M., and Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* 73, 1162–1169. doi: 10.1086/379378

Stojkovic, B., Sayadi, A., Dordevic, M., Jovic, J., Savkovic, U., and Arnqvist, G. (2017). Divergent evolution of life span associated with mitochondrial DNA evolution. *Evolution* 71, 160–166. doi: 10.1111/evo.13102

Suhara, H., and Kurogi, S. (2015). *Cantharellus cyphelloides* (Cantharellales), a new and unusual species from a Japanese evergreen broad-leaved forest. *Mycol. Prog.* 14:55. doi: 10.1007/s11557-015-1079-2

Sun, S., Fu, C., Ianiri, G., and Heitman, J. (2020). The pheromone and pheromone receptor mating-type locus is involved in controlling uniparental mitochondrial inheritance in cryptococcus. *Genetics* 214, 703–717. doi: 10.1534/genetics.119.302824

Thorn, R. G., Kim, J. I., Lebeuf, R., and Voitk, A. (2017). Golden chanterelles of newfoundland and labrador: a new species, a new record for north america, and a lost species rediscovered. *Botany* 95, 547–560. doi: 10.1139/cjb-2016-0213

Tian, X. F., Buyck, B., Shao, S. C., Liu, P. G., and Fang, Y. (2012). Cantharellus zangii, a new subalpine basidiomycete from southwestern China. *Mycotaxon* 120, 99–103. doi: 10.5248/120.99

Verma, S., Shakya, V. P. S., and Idnurm, A. (2018). Exploring and exploiting the connection between mitochondria and the virulence of human pathogenic fungi. *Virulence* 9, 426–446. doi: 10.1080/21505594.2017.1414133

Vielba-Fernandez, A., Bellon-Gomez, D., Tores, J. A., De Vicente, A., Perez-Garcia, A., and Fernandez-Ortuno, D. (2018). Heteroplasmy for the cytochrome b gene in *Podosphaera xanthii* and its role in resistance to QoI fungicides in spain. *Plant Dis.* 102, 1599–1605. doi: 10.1094/PDIS-12-17-1987-RE

Wang, L., Zhao, X., Li, X., and Li, L. (2009). Research status of the *Cantharellus* Population. *Edible Fungi of China* 28, 6–8.

Wang, P. F., Sha, T., Zhang, Y. R., Cao, Y., Mi, F., Liu, C. L., et al. (2017). Frequent heteroplasmy and recombination in the mitochondrial genomes of the basidiomycete mushroom *Thelephora ganbajun*. *Sci. Rep.* 7:1626. doi: 10.1038/s41598-017-01823-z

Wang, Y., and Hall, I. R. (2004). Edible ectomycorrhizal mushrooms: challenges and achievements. *Can. J. Bot.* 82, 1063–1073. doi: 10.1139/B04-051

Wang, Y., and Xu, J. (2020). Mitochondrial genome polymorphisms in the human pathogenic fungus *Cryptococcus neoformans*. *Front. Microbiol.* 11:706. doi: 10.3389/fmicb.2020.00706

Watling, R. (1997). Mycology–the business of fructification. *Nature* 385, 299–300. doi: 10.1038/385299a0

Wilson, A. J., and Xu, J. (2012). Mitochondrial inheritance: diverse patterns and mechanisms with an emphasis on fungi. *Mycology* 3, 158–166. doi: 10.1080/21501203.2012.684361

Xu, J. (2005). The inheritance of organelle genes and genomes: patterns and mechanisms. *Genome* 48, 951–958. doi: 10.1139/g05-082

Xu, J. (2007). "Origin, evolution, and extinction of asexual fungi," in *Sex in fungi*, eds J. Heitman, J. W. Kronstad, J. W. Taylor, and L. A. Casselton (Washington, DC: ASM Press), 461–475.

Xu, J. (2016). Fungal DNA barcoding. *Genome* 59, 913–932. doi: 10.1139/gen-2016-0046

Xu, J. (2020). Fungal species concepts in the genomics era. *Genome* 63, 459–468. doi: 10.1139/gen-2020-0022

Xu, J., Horgen, P. A., and Anderson, J. B. (1996). Somatic recombination in the cultivated mushroom *Agaricus bisporus*. *Mycol. Res.* 100, 188–192. doi: 10.1016/S0953-7562(96)80119-5

Xu, J., and Li, H. (2015). Current perspectives on mitochondrial inheritance in fungi. *Cell Heal. Cytos.* 7, 143–154. doi: 10.2147/Chc.S59508

Xu, J., and Wang, P. F. (2015). Mitochondrial inheritance in basidiomycete fungi. *Fung. Biol. Rev.* 29, 209–219. doi: 10.1016/j.fbr.2015.02.001

Xu, J., Yan, Z., and Guo, H. (2009). Divergence, hybridization, and recombination in the mitochondrial genome of the human pathogenic yeast *Cryptococcus gattii*. *Mol. Ecol.* 18, 2628–2642. doi: 10.1111/j.1365-294X.2009.04227.x

Xu, J., Zhang, Y., and Pun, N. (2013). Mitochondrial recombination in natural populations of the button mushroom *Agaricus bisporus*. *Fung. Genet. Biol.* 55, 92–97. doi: 10.1016/j.fgb.2012.09.004

Xue, B. T., Guo, J. L., Que, Y. X., Fu, Z. W., Wu, L. G., and Xu, L. P. (2014). Selection of suitable endogenous reference genes for relative copy number detection in sugarcane. *Int. J. Mol. Sci.* 15, 8846–8862. doi: 10.3390/ijms15058846

Yan, Z., Sun, S., Shahid, M., and Xu, J. (2007). Environment factors can influence mitochondrial inheritance in the fungus *Cryptococcus neoformans*. *Fung. Genet. Biol.* 44, 315–322. doi: 10.1016/j.fgb.2006.10.002

Yan, Z., and Xu, J. (2003). Mitochondria are inherited from the MATa parent in crosses of the basidiomycete fungus *Cryptococcus neoformans*. *Genetics* 163, 1315–1325. doi: 10.1017/S0016672303006141

Yang, Z. (2002). On wild mushroom resources and their utilization in Yunnan Province, Southwest China. *J. Nat. Resour.* 4, 463–469. doi: 10.1088/1009-1963/11/5/313

Zhang, Y., Mo, M., Yang, L., Mi, F., Cao, Y., Liu, C., et al. (2021). Exploring the species diversity of edible mushrooms in Yunnan, Southwestern China, by DNA barcoding. *J. Fungi* 7:310. doi: 10.3390/jof7040310

Zhang, Y., Yang, G. Z., Fang, M. L., Deng, C., Zhang, K. Q., Yu, Z. F., et al. (2020). Comparative analyses of mitochondrial genomes provide evolutionary insights into nematode-trapping fungi. *Front. Microbiol.* 11:617. doi: 10.3389/fmicb.2020.00617

Zhou, J. W., Liu, L. M., and Chen, J. (2010). Mitochondrial DNA heteroplasmy in *Candida glabrata* after mitochondrial transformation. *Eukary. Cell* 9, 806–814. doi: 10.1128/EC.00349-09

Conflict of Interest: The 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fmicb.2021.707281

Comparative Mitogenomic Analysis and the Evolution of *Rhizoctonia solani* Anastomosis Groups

Runmao Lin[1†], Yuan Xia[2†], Yao Liu[3†], Danhua Zhang[2†], Xing Xiang[2], Xianyu Niu[2], Linjia Jiang[2], Xiaolin Wang[2] and Aiping Zheng[2,4]*

[1] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Agriculture College, Sichuan Agricultural University, Chengdu, China, [3] Rice Research Institute, Sichuan Agricultural University, Chengdu, China, [4] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Chengdu, China

Mitochondria are the major energy source for cell functions. However, for the plant fungal pathogens, mitogenome variations and their roles during the host infection processes remain largely unknown. *Rhizoctonia solani*, an important soil-borne pathogen, forms different anastomosis groups (AGs) and adapts to a broad range of hosts in nature. Here, we reported three complete mitogenomes of AG1-IA RSIA1, AG1-IB RSIB1, and AG1-IC, and performed a comparative analysis with nine published *Rhizoctonia* mitogenomes (AG1-IA XN, AG1-IB 7/3/14, AG3, AG4, and five *Rhizoctonia* sp. mitogenomes). These mitogenomes encoded 15 typical proteins (*cox1-3*, *cob*, *atp6*, *atp8-9*, *nad1-6*, *nad4L*, and *rps3*) and several LAGLIDADG/GIY-YIG endonucleases with sizes ranging from 109,017 bp (*Rhizoctonia* sp. SM) to 235,849 bp (AG3). We found that their large sizes were mainly contributed by repeat sequences and genes encoding endonucleases. We identified the complete sequence of the *rps3* gene in 10 *Rhizoctonia* mitogenomes, which contained 14 positively selected sites. Moreover, we inferred a robust maximum-likelihood phylogeny of 32 Basidiomycota mitogenomes, representing that seven *R. solani* and other five *Rhizoctonia* sp. lineages formed two parallel branches in Agaricomycotina. The comparative analysis showed that mitogenomes of Basidiomycota pathogens had high GC content and mitogenomes of *R. solani* had high repeat content. Compared to other strains, the AG1-IC strain had low substitution rates, which may affect its mitochondrial phylogenetic placement in the *R. solani* clade. Additionally, with the published RNA-seq data, we investigated gene expression patterns from different AGs during host infection stages. The expressed genes from AG1-IA (host: rice) and AG3 (host: potato) mainly formed four groups by k-mean partitioning analysis. However, conserved genes represented varied expression patterns, and only the patterns of *rps3-nad2* and *nad1-m3g18/mag28* (an LAGLIDADG endonuclease) were conserved in AG1-IA and AG3 as shown by the correlation coefficient analysis, suggesting regulation of gene repertoires adapting to infect varied hosts. The results of variations in mitogenome characteristics and the gene substitution rates and expression patterns may provide insights into the evolution of *R. solani* mitogenomes.

Keywords: *Rhizoctonia solani*, mitogenome, expression pattern, evolution, positive selection

OPEN ACCESS

Edited by:
Georg Hausner, University of Manitoba, Canada

Reviewed by:
Hilal Ozkilinc, Çanakkale Onsekiz Mart University, Turkey
James Hane, Curtin University, Australia
Georgios Tzelepis, Swedish University of Agricultural Sciences, Sweden

***Correspondence:**
Aiping Zheng
apzh0602@gmail.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *09 May 2021*
Accepted: *30 August 2021*
Published: *20 September 2021*

Citation:
Lin R, Xia Y, Liu Y, Zhang D, Xiang X, Niu X, Jiang L, Wang X and Zheng A (2021) Comparative Mitogenomic Analysis and the Evolution of Rhizoctonia solani Anastomosis Groups. Front. Microbiol. 12:707281. doi: 10.3389/fmicb.2021.707281

INTRODUCTION

The basidiomycetous fungus *Rhizoctonia solani* Kühn [teleomorph *Thanatephorus cucumeris* (Frank) Donk] is a worldwide prevalent soil-borne plant pathogen. It causes diseases in many economically important crops (including rice, corn, soybeans, potatoes, wheat, cabbage, lettuce, sugar beets, and tomatoes), ornamental plants, and forest trees (Ogoshi, 1987; Gonzalez Garcia et al., 2006; Yang and Li, 2012; Molla et al., 2020).

The multinucleate *R. solani* isolates are grouped in the taxa within the *Rhizoctonia* species complex (Carling, 1996). These *R. solani* isolates are classified into 14 distinct anastomosis groups (i.e., AG1-AG13 and AGBI), and AG1 consists of four primary intraspecific subgroups of AG1-IA, AG1-IB, AG1-IC, and AG1-ID (Pannecoucque and Höfte, 2009; Yang and Li, 2012). Among AG1-AG13, strains are generally capable of fusing hyphae only in each AG, while strains from AGBI are capable of fusing hyphae with strains from 14 AGs (Sneh et al., 1991; Gonzalez Garcia et al., 2006). Meanwhile, the binucleate *Rhizoctonia* sp. isolates are classified in other taxa within the complex (Carling, 1996), which include 21 AGs (AG A-U).

In recent years, *R. solani* pathogenesis has been studied at the genomic and transcriptomic level, including AG1-IA that causes sheath blight in rice (*Oryza sativa*), corn (*Zea mays*), and soybeans (*Gly*cine max; Zheng et al., 2013; Nadarajah et al., 2017; Xia et al., 2017; Yamamoto et al., 2019; Lee et al., 2021; Li et al., 2021a), AG1-IB that infects lettuce (Wibberg et al., 2015; Verwaaijen et al., 2017), AG3 that infects potatoes (Cubeta et al., 2014; Patil et al., 2018; Zrenner et al., 2020), and AG8 that infects wheat (Hane et al., 2014).

Mitochondrial genomes evolve independently of the nuclear genomes, and comparative mitogenome analysis sheds light on mitochondrial evolution (Gray, 2012). The relatively small size and mostly uniparental inheritance of fungal mitochondria also makes them ideal candidates for evolution, fungicide insensitivity, population genetics, and taxonomy studies (Bullerwell and Lang, 2005). So far, more than 800 complete fungal mitogenomes are available in the NCBI database[1], providing a rich picture of their prevailing features, ancestral characteristics, and evolutionary trends. About 16% of these mitogenomes are in Basidiomycetes, including AG1-IB 7/3/14 and AG3 mitogenomes (Wibberg et al., 2013; Losada et al., 2014). The partial mitogenomes of AG1-IA and AG8 have also been reported (Zheng et al., 2013; Hane et al., 2014). The AG3 mitogenome revealed the expansion of mobile elements in *R. solani* and the synteny among AG1-IA, AG1-IB, and AG3 mitogenomes (Losada et al., 2014). Recently, multi-, bi-, and uninucleate *Rhizoctonia* mitogenomes have been reported (Li et al., 2021b). However, lack of complete mitogenomes of AG1-IA and AG1-IC that can infect cabbage and soybeans (Fu et al., 2014; Misawa and Aoki, 2017) prevents our understanding of the diversity of mitogenomic characteristics in *R. solani*. A comparison of mitogenomes in Basidiomycota could provide valuable insight into the origin and evolution of their complex mitogenomic features.

In fungal biology, mitochondria play a significant role in fungal virulence and adaptation (Ingavale et al., 2008; Chatre and Ricchetti, 2014; Calderone et al., 2015; Sun et al., 2019). Previous studies show that mutations in the mitogenome of the tree pathogen *Cryphonectria parasitica* weaken its virulence (Monteiro-Vitorello et al., 1995), and the mitochondrial cytochrome C from the animal pathogen *Aspergillus fumigatus* is critical for its virulence (Grahl et al., 2012). For the human opportunistic pathogen *Cryptococcus neoformans*, the changes in its mitochondria morphology by fission and fusion could dramatically influence its virulence (Chang and Doering, 2018). Meanwhile, the mitochondria of *C. neoformans* play a key role in hypoxia adaptation (Ingavale et al., 2008). Moreover, lineage-specific adaptations in mitochondria have been found to be associated with hosts in another opportunistic pathogen, *Candida albicans*, and the mitochondrial proteins influence *C. albicans* respiration (Sun et al., 2019) that is required for its growth, morphogenesis, and virulence. Many chemicals can efficiently inhibit respiration in *C. albicans* while not damaging the mammalian host (Duvenage et al., 2019), which may be a strategy to develop a target for antifungals in the future studies.

The fungal mitogenomes may be a powerful system to measure adaptation at the molecular level. The estimation of substitution rates may provide evidence of adaptive evolution that possibly affects only a few amino acids at a few time points (Yang and Nielsen, 2002). To measure the selection pressure on amino acid replacement mutation of protein-coding genes, the method of calculating the non-synonymous/synonymous substitution rate ratio (dN/dS) is widely used (Nielsen and Yang, 2003). Based on eight mitogenomes of the *Synalpheus* species of non-eusocial and eusocial groups, the comparative analyses of synonymous substitution rates and selection signals provide direct evidence of eusociality impacting genome evolution (Chak et al., 2021). The discovery of several positively selected sites in eusocial lineages may represent adaptation (Chak et al., 2021). For host specificity of *R. solani* AG strains, the examination of substitution rates in mitogenomes may help to reveal their adaptability to hosts.

Additionally, the expression of *R. solani* nuclear genes during host infection enhances our discovery of pathogenic factors, including candidate effectors (Zheng et al., 2013). The interactions between AG1-IA and rice, AG1-IB and lettuce, as well as AG3 and potatoes have been explored (Zheng et al., 2013; Verwaaijen et al., 2017; Xia et al., 2017; Zrenner et al., 2020), providing an avenue to investigate the expression of mitochondrial encoded genes and their roles during infection process. A schematic of varied gene expression patterns during infecting different hosts may provide clues to understand fungal adaptation to hosts.

Here, in exploring the evolution and host adaptation in *R. solani* by performing comparative analysis of mitogenomes, we reported three complete mitogenomes of *R. solani* AG1-IA RSIA1, AG1-IB RSIB1, and AG1-IC, providing a resource for revealing mitogenome characteristics. We also investigated the phylogenetic analysis and selection pressure analysis on amino acids, which may indicate significant sites contributing

[1]https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/; April 2021

to adaptation, and examined the varied expression patterns of encoded genes from mitogenomes of AG strains with host specificity during infection, which may further provide knowledge about host adaptation in *R. solani*.

MATERIALS AND METHODS

Sampling and DNA Extraction

The *R. solani* AG1-IA RSIA1, AG1-IB RSIB1, and AG1-IC strains were provided by Prof. Erxun Zhou at South China Agricultural University. The strains were grown in potato dextrose broth medium at 28°C, and the genomic DNA was extracted using a modified CTAB method (Ciampi et al., 2008). All the *Rhizoctonia* mitogenomes used in this study were listed in **Supplementary Table 1**, including the previously reported mitogenomes of AG1-IB 7/3/14 (Wibberg et al., 2013), AG3 (Losada et al., 2014), AG4 (Zhang et al., 2021), AG1-IA XN, *R. cerealis* RW, and four *Rhizoctonia* sp. strains, JN, LY, SM, and YR (Li et al., 2021b). The hosts of five complete mitogenomes were listed in **Table 1**.

Mitogenome Assembly and Annotation

For the sequenced PacBio RS long reads of AG1-IA RSIA1, AG1-IB RSIB1, and AG1-IC strains, we used LoRDEC v0.5 (Salmela and Rivals, 2014) with parameters of "-k 19 -s 3" for read correction based on Illumina short reads with insert size of ∼180 bp. Then we used Canu v1.2 (Koren et al., 2017) with default parameters for genome assembly, which generated the complete three mitogenomes of AG1-IA, AG1-IB, and AG1-IC. We examined the circular map of the mitogenomes and improved the sequences using Pilon v1.17 (Walker et al., 2014) with default parameters.

From the mitogenome sequences, we predicted and annotated the 15 typical protein-coding genes (seven subunits of NADH dehydrogenase, three cytochrome c oxidase subunits, three ATP synthase subunits, one apocytochrome b, and one ribosomal protein) and other protein-coding genes (LAGLIDADG homing endonucleases and GIY-YIG endonucleases) by the pipeline as follows. First, we aligned the mitogenome sequences against amino acids in the NCBI NR database using BLASTPX with an *E*-value cutoff of 1e-10, which detected candidate reference genes from the NR database. Then, we used Exonerate v2.2.0 (Slater and Birney, 2005) with the "protein2genome" model to predict genes by aligning mitogenome sequences against these candidate reference genes. We found that some Exonerate-predicted genes may be incomplete without considering the start and/or stop codons. For each predicted gene, we wrote an in-house Perl script to check and improve the prediction by scanning its up-/down-stream genomic sequences to identify the start and stop codons. For each gene region, Exonerate may predict multiple candidate genes because of multiple NR reference genes being used for alignment analysis. All candidate genes were aligned to NR reference genes again using BLASTP, which could be useful for manual examination of the length and *E*-value for each predicted gene. For multiple predicted genes from the same genomic region, we manually selected the one with the low BLASTP *E*-value and with similar length compared to the NR reference genes. Finally, the annotation of selected genes was inferred from NR reference genes. For tRNA genes, we used tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997) with translation table 4 for gene discovery and removed candidate tRNAs with types of "Undet" (i.e., without anticodons). The reported rRNA sequences in the SILVA database (Quast et al., 2013) were used as reference genes for rRNA annotation by performing BLASTN analysis.

We used the same method to annotate the encoding genes of the previously reported AG1-IB 7/3/14, AG3, AG4, AG1-IA XN, *R. cerealis* RW, and four *Rhizoctonia* sp. mitogenome sequences (JN, LY, SM, and YR). The AG4 mitogenome was included in the reported sequence deposited in NCBI with the accession number of JADHEA010000014.1 (Zhang et al., 2021;

TABLE 1 | Statistics of mitochondrial genomes of *Rhizoctonia solani* species.

	AG1-IA RSIA1	AG1-IB RSIB1	AG1-IB 7/3/14	AG1-IC	AG3-PT
Complete genome	Yes	Yes	Yes	Yes	Yes
Genome size (bp)	152,549	168,442	162,751	175,227	235,859
Typical proteins[a]	15	15	15	15	15
Other proteins[b]	36	12	15	24	21
tRNA	26	26	25	26	27
rRNA	2	2	2	2	2
GC content (%)	33.7	36.5	36.4	34.5	35.9
Repeat sequences (%)	17.86	32.17	30.52	28.96	42.73
Accession	MW995474.1	MW995475.1	HF546977.1	MW995476.1	KC352446.1
Reference	This study	This study	Wibberg et al., 2013	This study	Losada et al., 2014
Host[c]	Rice, corn, soybeans, barley, sorghum, potatoes, peanut, cabbage, leaf lettuce, et al.	Corn, sugar beets, gay feather, common bean, soybeans, cabbage, leaf lettuce et al.		Sugar beets, carrot, buckwheat, flax, soybeans, bean, cabbage, pineapple, panicum, spinach and radish	Potatoes with black scurf symptoms

[a] *The 15 genes include cytochrome c oxidase subunits (cox1, cox2, and cox3), apocytochrome b (cob), ATP synthase subunits (atp6, atp8, and atp9), subunits of NADH dehydrogenase (nad1, nad2, nad3, nad4, nad4L, nad5, and nad6), and one ribosomal protein (rps3).*
[b] *LAGLIDADG homing endonucleases and GIY-YIG endonucleases.*
[c] *The host information was inferred from study by Yang and Li (2012).*

Supplementary Table 2). We found that there was one base deletion in the *rps3* gene (**Supplementary Figure 1**), preventing the prediction of the *rps3* gene in AG4 (**Supplementary Table 3**). As we could not be sure that the AG4 *rps3* was a real pseudogene or had an error in assembly sequence, we did not include AG4 *rps3* for comparative analysis. For the other six mitogenomes, we found that there were 17, 4, 1, 1, and 1 gap regions (i.e., "Ns" in assemblies) in XN (accession number: MT887631.1), LY (accession number: MT887629.1), SM (accession number: MT887628.1), YR (accession number: MT887627.1), and RW (accession number: MT887630.1) mitogenomes, respectively. The *Rhizoctonia* sp. JN (accession number: MT887626.1) mitogenome did not have a gap sequence, but its length (∼126 kb) was ∼35 kb less than that of *Rhizoctonia* sp. LY (∼161 kb), preventing the confirmation of complete mitogenome of *Rhizoctonia* sp. JN. The incomplete mitogenomes may prevent the prediction of genes (such as the incomplete *rps3* in RW strain). However, the complete sequences of 14 typical proteins (*cox1-3*, *cob*, *atp6*, *atp8-9*, *nad1-6,* and *nad4L*) were identified in 12 *Rhizoctonia* mitogenomes. To perform comparative analysis of endonucleases, only five complete mitogenomes of AG1-IA RSIA1, AG1-IB RSIB1, AG1-IC, AG1-IB 7/3/14, and AG3 were used (**Table 1** and **Supplementary Table 4**).

We used the *de novo* method to identify repeat sequences in *Rhizoctonia* mitogenomes. The repeat library was constructed based on the mitochondrial genome sequences using RepeatScout v1.0.5 (Price et al., 2005). This library was used to identify repeat sequences using RepeatMasker v4.0.5[2].

Comparative Mitogenomic Analysis

Based on amino acid sequences of genes from five complete mitogenomes of *R. solani* strains, we used OrthoFinder 0.7.1 (Emms and Kelly, 2015) to detect their orthologous genes. The sequence alignment of the *rps3* gene was done by MUSCLE v3.8.31 (Edgar, 2004). The positively selected signals in *rps3* genes were detected using CODEML implemented in PAML v 4.8a (Yang, 2007), as described in the previous study (Lin et al., 2015). For the *rps3* gene with positively selected signals, we used PSIPRED (Buchan and Jones, 2019) and RoseTTAFold (Baek et al., 2021) to predict its protein structure.

The KaKs_Calculator 1.2 estimated dN and dS values using model-selected and model-averaged methods based on a group of models (Zhang et al., 2006). As in the description in the KaKs_Calculator study (Zhang et al., 2006), different substitution models considered different evolutionary features, resulting in different estimates, and for protein-coding sequences, the use of many features may lead to more reliable evolutionary information. We used the 10 methods (NG, LWL, MLWL, LPB, MLPB, GY-HKY, YN, MYN, MS, and MA) implemented in the KaKs_Calculator to estimate dN, dS, and dN/dS values for protein-coding genes in *Rhizoctonia* mitogenomes. We used their mean values to represent the increasing or decreasing trends of the dN, dS, and dN/dS values for the comparative analysis.

To perform phylogeny analysis for 32 mitogenomes from Basidiomycetes (**Supplementary Table 1**), we selected amino acids from 14 typical protein-coding genes (*cox1*, *cox2*, *cox3*, *cob*, *atp6*, *atp8*, *atp9*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5,* and *nad6*) and performed MUSCLE alignment. Then these sequences were concatenated for the following analysis. The ProtTest v3.4 (Darriba et al., 2011) with parameters of "-all-distributions -F -AIC -BIC" identified the best model of LG + I + G + F for constructing the maximum-likelihood phylogeny. Then we used Mega v6.06 (Tamura et al., 2013) to build the maximum-likelihood phylogenetic tree with bootstrap value of 1,000.

Transcriptomic Analysis

For AG1-IA, its gene expression analysis was investigated using RNA-seq after rice infection at 10 h (10-h), 18, 24, 32, 48, and 72-h (Zheng et al., 2013). The RNA-seq data before and after infecting different crops (i.e., rice, corn, and soybeans) of different AG1-IA strains that were isolated from rice, corn, and soybeans have been reported (Xia et al., 2017). For three strains of AG3, their interaction with potato sprouts after infection of three and 8 days were investigated by transcriptomic analysis (Zrenner et al., 2020). For the reported RNA-seq data, we analyzed data from each study independently. We calculated the gene expression FPKM (fragments per kilo base per million mapped reads) values following the protocol (Pertea et al., 2016) using HISAT2 (Kim et al., 2019), StringTie (Pertea et al., 2015), and Ballgown (Frazee et al., 2015) software. Based on the expression values, we used the R function of fviz_cluster that was implemented in the factoextra package to detect gene clusters and used the R function cor to calculate the Pearson correlation coefficient between genes.

Gene Expression Analysis via Real-Time Quantitative Reverse Transcription PCR (RT-qPCR)

A total of 15 genes were selected for RT-qPCR analysis. First-strand cDNA was synthesized from total RNA using HiScript II Q RT Supermix for qPCR with a gDNA wiper (Vazyme R223-01, Nanjing, China). RT-qPCR was performed using the AceQ qPCR SYBR green master mix (Vazyme Q111-02/03, Nanjing, China). The qPCR reactions were performed in a final volume of 10 μL containing 5 μL of 2 × AceQ qPCR SYBR green master mix, 0.25 μL of 10 μM of each primer, 4.25 μL of ddH_20, and 0.25 μL of cDNA. Reactions were carried out at 95°C for 5 min, followed by 40 cycles of 95°C for 10 s, 60°C for 30 s, and melting curve analysis from 60°C to 95°C at 1°C increments (qTOWER^{3}G, Jena German). Primers for qPCR were designed based on our predicted gene sequences by NCBI primer blast, and the parameters were modified to suit the RT-qPCR conditions (**Supplementary Table 5**). The 18S gene was used as an internal control. Fold changes were determined by the $2^{-\Delta\Delta Ct}$ method. All qPCR reactions were run in triplicate.

RESULTS

General Characteristics of *R. solani* Mitogenomes

Here we reported three complete mitogenomes of *R. solani* AG1-IA, AG1-IB, and AG1-IC, with the sizes of ∼152-168 kb (**Table 1**), and performed a comparative analysis with two

[2]http://www.repeatmasker.org/

published complete mitogenomes of *R. solani* AG1-IB and AG3 (Wibberg et al., 2013; Losada et al., 2014). Among the five complete mitogenomes, the smallest size was ∼152 kb in AG1-IA and the largest size was ∼235 kb in AG3 (**Table 1**). We found highly conserved sequences in the two AG1-IB mitogenomes of RSIB1 and 7/3/14 strains. These mitogenomes consisted of an essential set of 15 typical protein-coding genes (three cytochrome c oxidase subunits: *cox1*, *cox2*, *cox3*; the apocytochrome b: *cob*; three ATP synthase subunits: *atp6*, *atp8*, *atp9*; seven subunits of NADH dehydrogenase: *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*; and one ribosomal protein: *rps3*), LAGLIDADG homing endonucleases and GIY-YIG endonucleases (ranging from 12 in AG1-IB and 36 in AG1-IA), and the small and large ribosomal RNA subunits (*rns*, *rnl*), and tRNAs (**Table 1**, **Figure 1**, and **Supplementary Table 4**). All protein-coding genes were clustered into 15, 14, 3, and 1 orthologous groups for 15 typical protein-coding genes, LAGLIDADG homing endonucleases, GIY-YIG endonucleases and DNA polymerase, respectively (**Figure 1B**). Most groups contained single-copy genes from each mitogenome, excluding three LAGLIDADG groups and one GIY-YIG group that each contained multiple-copy genes (**Figure 1C**). For example, the RSMOG01 group contained only one *cox1* in each strain, while the RSMOG16 group contained LAGLIDADG homing endonucleases ranging from 3 in AG1-IB 7/3/14 to 18 in AG1-IA. Compared to other strains, AG1-IA contained more LAGLIDADG homing endonucleases that were mainly encoded in the intron regions of *rnl*, *cox1,* and *nad4L* (**Figure 1A** and **Supplementary Table 4**).

Among these mitogenomes, there were 27,239 (17.86%), 54,190 (32.17%), 49,669 (30.52%), 50,748 (28.96%), and 100,785 (42.73%) bp of repeat sequences in AG1-IA, AG1-IB, AG1-IB 7/3/14, AG-IC, and AG3, respectively (**Supplementary Tables 6–10**), with the lowest and highest ratios in AG1-IA and AG3, respectively. The genomic size of AG3 was 83,310 bp larger than that of AG1-IA (**Table 1**), while the repeat sequences of AG3 were 73,546 bp larger than those of AG1-IA, indicating that repetitive sequences contribute to the large size of the AG3 mitogenome. The comparison of genomic sizes and ratios of repeat sequences suggested their positive correlations ($R = 0.81$, $P = 0.0015$), i.e., longer genomic sizes containing more repetitive sequences (**Figure 1D**).

However, the relationship between GC contents and genomic sizes was not similar to that between repeat sequences and genomic sizes (**Figure 1E**). Although AG3 had the largest genomic size, its GC content was larger than 33.7% of the GC content in AG1-IA and was lower than 36.5% for AG-IB (**Table 1**). The distribution of GC content among mitogenomes may suggest the different sequence preferences in mitogenomes.

Mitochondrial Phylogenetic Relationships Between *R. solani* and Other Fungi in Basidiomycota

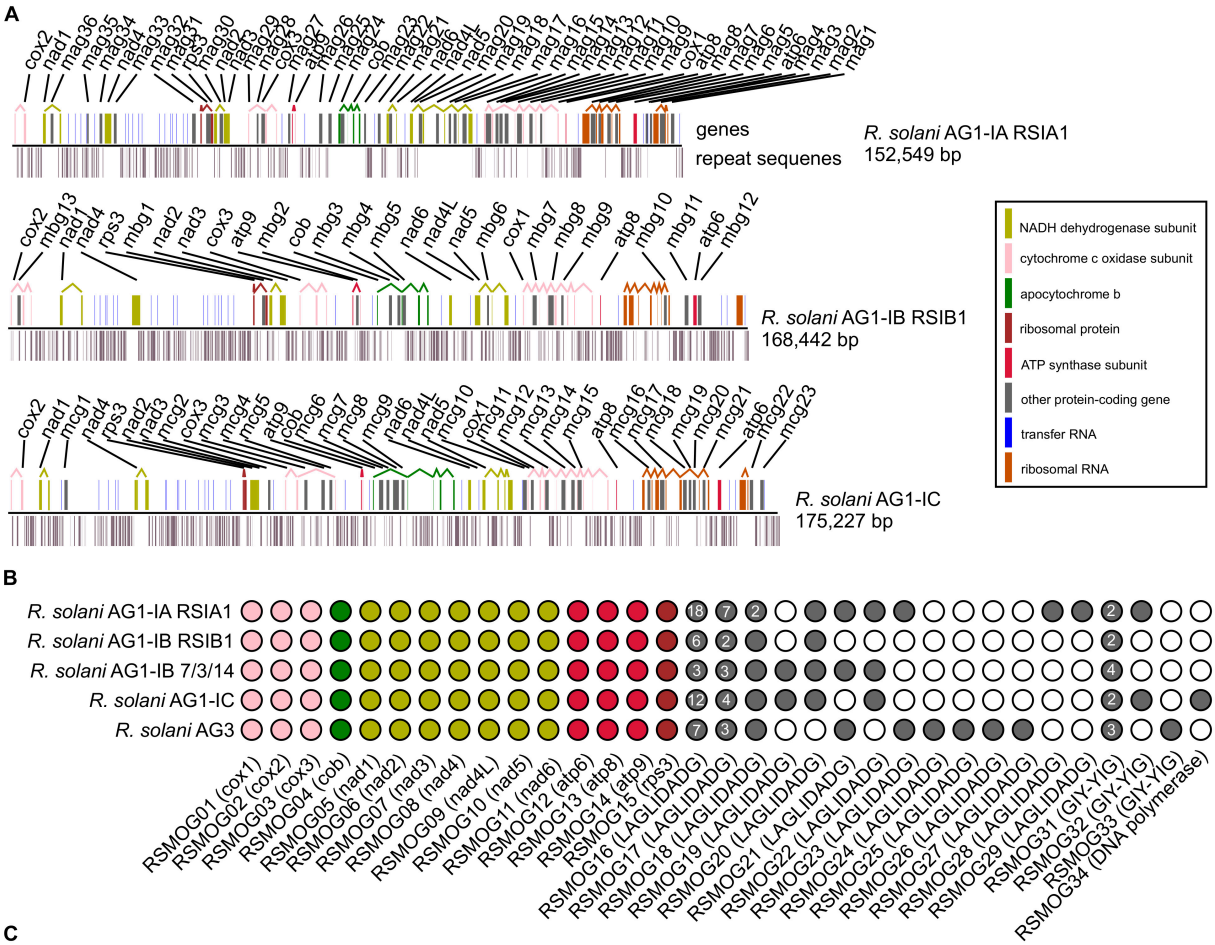
Based on the complete mitogenome of *R. solani*, we explored their phylogenetic relationships with other fungi. A phylogeny for 32 fungal strains in Basidiomycota and one strain in Ascomycota as an outgroup was constructed, which represented 26, 2, and 3 Basidiomycetes strains in three subphyla of Agaricomycotina, Pucciniomycotina, and Ustilaginomycotina, respectively (**Figure 2A**). *Rhizoctonia* strains were in Agaricomycotina. The seven multinucleate *R. solani* strains (AG1-IA RSIA1 and XN, AG1-IB RSIB1 and 7/3/14, AG1-IC, AG3, and AG4) were parallel with a clade containing two binucleate *Rhizoctonia* strains (*Rhizoctonia* sp. LY and *R. cerealis* RW) and three uninucleate *Rhizoctonia* strains (*Rhizoctonia* sp. SM, JN, and YR; Li et al., 2021b). These *Rhizoctonia* lineages were in the Cantharellales order, plus the *Cantharellus cibarius* lineage formed one large clade that was parallel with another clade containing 12 lineages in four orders (Agaricales, Corticiales, Russulales, and Polyporales). Outside the branches of these 25 lineages, there was one branch for *Serendipita indica* in the Sebacinales order (in Agaricomycotina).

From the phylogeny, mitochondrial genomic sizes varied from 29 kb (*Jaminaea angkorensis* strain; in Microstromatales, Ustilaginomycotina) to 235 kb (AG3; in Cantharellales, Agaricomycotina). We found that five strains from Pucciniomycotina and Ustilaginomycotina had mitochondrial genomic sizes of less than 41 kb (**Supplementary Table 1**), excepting the *Sporisorium reilianum* strain of ∼90 kb in size (in Ustilaginales, Ustilaginomycotina). However, the mitochondrial genomic sizes were obviously increased in strains from Agaricomycotina, a separate clade in Basidiomycota, especially for *R. solani* strains in Cantharellales and *Phlebia radiata* strain in Corticiales, with sizes larger than 150 kb (**Figure 2** and **Supplementary Table 1**).

Considering both GC content and mitogenomic sizes, we found that in Pucciniomycotina and Ustilaginomycotina, most mitogenomes had small sizes but had high GC content (>31%). In Agaricomycotina, the GC content was quite different, ranging from 22.8 to 39.66% (**Figures 2B,C** and **Supplementary Table 1**). A positive relationship ($R = 0.73$, $P = 2.7e\text{-}05$) between GC content and mitogenomic sizes are shown for Agaricomycotina strains, i.e., strains with higher genomic sizes with higher GC content (**Figure 2C**). Meanwhile, the repeat sequences in the mitogenomes had little effect on GC content (**Figure 2D**). Moreover, in Basidiomycota fungi, high mitochondrial GC content was found in pathogens (including *Phakopsora* sp. in Pucciniomycotina, *Malassezia* sp., and *Sporisorium* sp. in Ustilaginomycotina, and *Rhizoctonia* sp. in Agaricomycotina; **Figure 2** and **Supplementary Table 1**).

Changes in Non-Synonymous and Synonymous Substitution Rates (dN and dS) of *R. solani* Mitogenomes

Although the *R. solani* phylogeny formed one branch in the mitochondrial phylogeny (**Figure 2A**), the AG1-IC and other AG1 strains were separated by the AG3 strain, which may reflect the sequence changes in mitogenomes. We used the KaKs_Calculator to calculate the dN and dS values for the concatenated sequences of 15 typical protein-coding genes and found that all dN, dS, and dN/dS values were lower than 0.03000, 0.25000, and 0.18000, respectively (**Figure 3** and **Supplementary Table 11**). For each pair of mitogenomes, the AG1-IC and AG3



GroupID	AG1-IA RSIA1	AG1-IB RSIB1	AG1-IB 7/3/14	AG1-IC	AG3	Annotation
RSMOG16	mag9 mag10 mag12 mag13 mag14 mag17 mag19 mag20 mag22 mag23 mag25 mag26 mag28 mag30 mag32 mag33 mag34 mag36	mbg3 mbg4 mbg5 mbg6 mbg11 mbg12	mbGg7 mbGg13 mbGg14	mcg2 mcg3 mcg4 mcg5 mcg7 mcg9 mcg10 mcg11 mcg14 mcg16 mcg19 mcg20	m3g4 m3g10 m3g11 m3g12 m3g14 m3g15 m3g18	LAGLIDADG homing endonuclease
RSMOG17	mag2 mag8 mag15 mag16 mag21 mag24 mag31	mbg1 mbg8	mbGg2 mbGg10 mbGg15	mcg6 mcg8 mcg13 mcg17	m3g2 m3g16 m3g19	LAGLIDADG homing endonuclease
RSMOG18	mag11 mag35	mbg9	mbGg3	mcg12	m3g20	LAGLIDADG homing endonuclease
RSMOG31	mag1 mag3	mbg2 mbg13	mbGg8 mbGg9 mbGg11 mbGg12	mcg22 mcg23	m3g6 m3g7 m3g8	GIY-YIG endonuclease

FIGURE 1 | Characterization of mitogenomes in *Rhizoctonia solani*. **(A)** Maps of six mitogenomes. The genomic positions of genes and repeat sequences were shown. Many protein-coding genes contained multiple exons that were connected by broken lines. **(B)** Orthologous groups of protein-coding genes in mitogenomes. **(C)** Four groups contained multiple genes. **(D,E)** Distribution of repeat sequences and GC contents in *Rhizoctonia* mitogenomes. In the MATERIALS AND METHODS section (Mitogenome assembly and annotation), we described selecting *Rhizoctonia* mitogenomes for comparative analysis. The R functions (ggplot and stat_cor with the Pearson correlation coefficient method implemented in ggplot2 and ggpubr packages, respectively) were used to draw the **(D,E)**.

mitogenomes had the highest dN values (i.e., 0.02339∼0.02934, mean 0.02583), while AG1-IB and AG3 had the lowest dN values (i.e., 0.01418∼0.01735, mean 0.01536). Similar results were found for dS values, i.e., the highest dS values were found for AG1-IC vs AG3 and the lowest dS values were found for AG1-IB vs AG3. The mitogenome pairs of AG1-IC and other *R. solani* mitogenomes showed the higher substitution rates than those from mitogenome pairs without AG1-IC (**Figure 3** and **Supplementary Table 11**), supporting the phylogenetic topology for *R. solani* mitogenomes (**Figure 2A**).

Discovery of Positively Selected Sites in *Rhizoctonia rps3* Genes

For each of 15 typical genes, we calculated their dN/dS values and found that all of them were less than one, including the *rps3* genes (**Figures 4A,B**). The amino acid (aa) sequences of *rps3* in four strains (AG1-IB RSIB1 and 3/7/14, AG1-IC, and AG3) were 283 aa, and one more aa was found in AG1-IA strains (RSIA1 and XN), as well as 56 more aa were found in *Rhizoctonia* sp. strains (LY, SM, JN, and YR). The sequence alignment showed that they shared sequence identities larger than 89%, suggesting the conserved sequences in *Rhizoctonia rps3* genes. However, with the CODEML method in PAML (Yang, 2007), we detected 14 positively selected sites (**Figure 4C**). Among them, five sites ("RPHA" and "A" in AG1-IA) were closely linked with each other (aa position: 84–89), with one amino acid ("L" in AG1-IA) flanking these positively selected sites; meanwhile, for the 27 aa downstream of these sites, other positively selected sites were found, including "P," "I," and "NTT" in AG1-IA (**Figure 4C**). These changed sites represented five types of different sequences in AG1-IA, AG1-IB, AG1-IC/AG3, SM, and other strains (LY, JN, and YR), respectively, which were related to their phylogenic topology (**Figure 2A**). The secondary structure

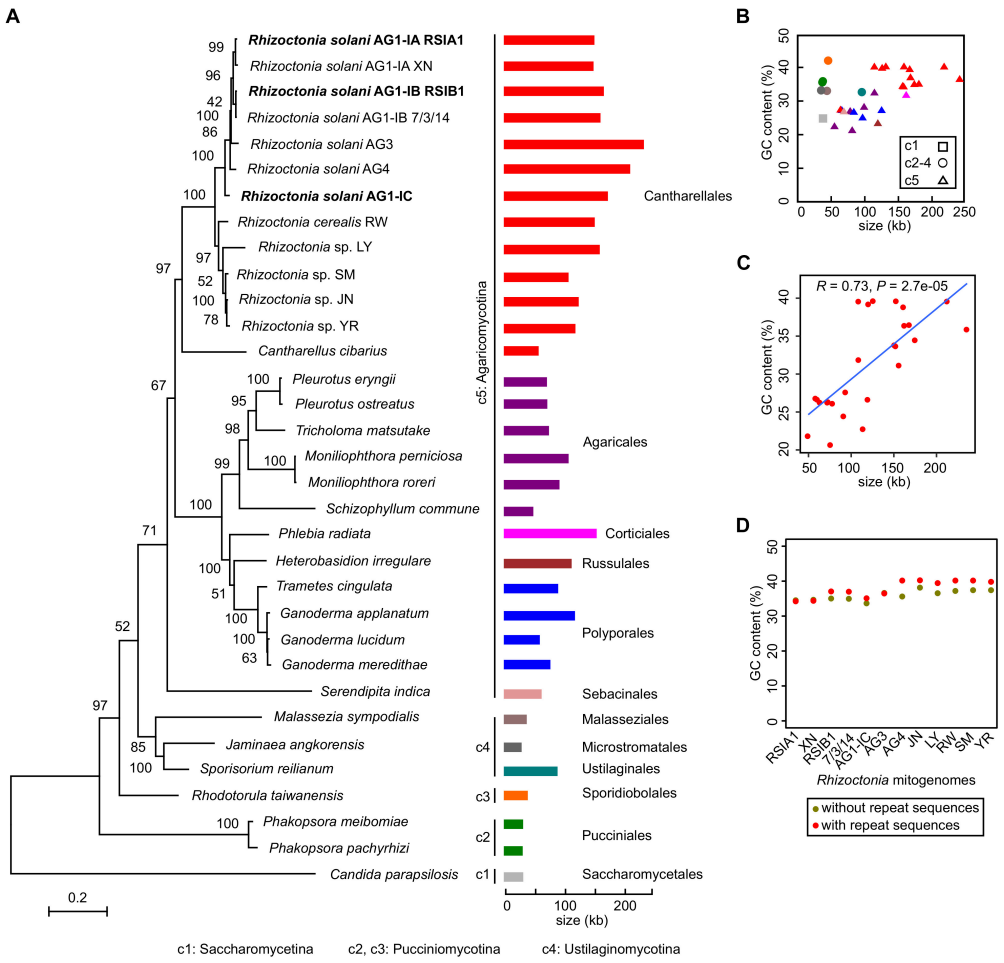


FIGURE 2 | Phylogenetic relationships between *Rhizoctonia solani* and other fungi in Basidiomycota. **(A)** The mitogenome phylogeny of 33 fungal strains. The *Candida parapsilosis* in Ascomycete was used as the outgroup. The maximum-likelihood bootstrap values were shown. The sizes of mitogenomes were shown with different colors for strains from different orders. The newly reported complete mitogenomes were marked in bold. **(B)** Distribution between sizes and GC content from 33 fungal mitogenomes. **(C)** Positive relationship between GC content and mitogenomic sizes for Agaricomycotina mitogenomes. **(D)** Distribution of GC content in *Rhizoctonia* mitogenomes with and without repeat sequences.

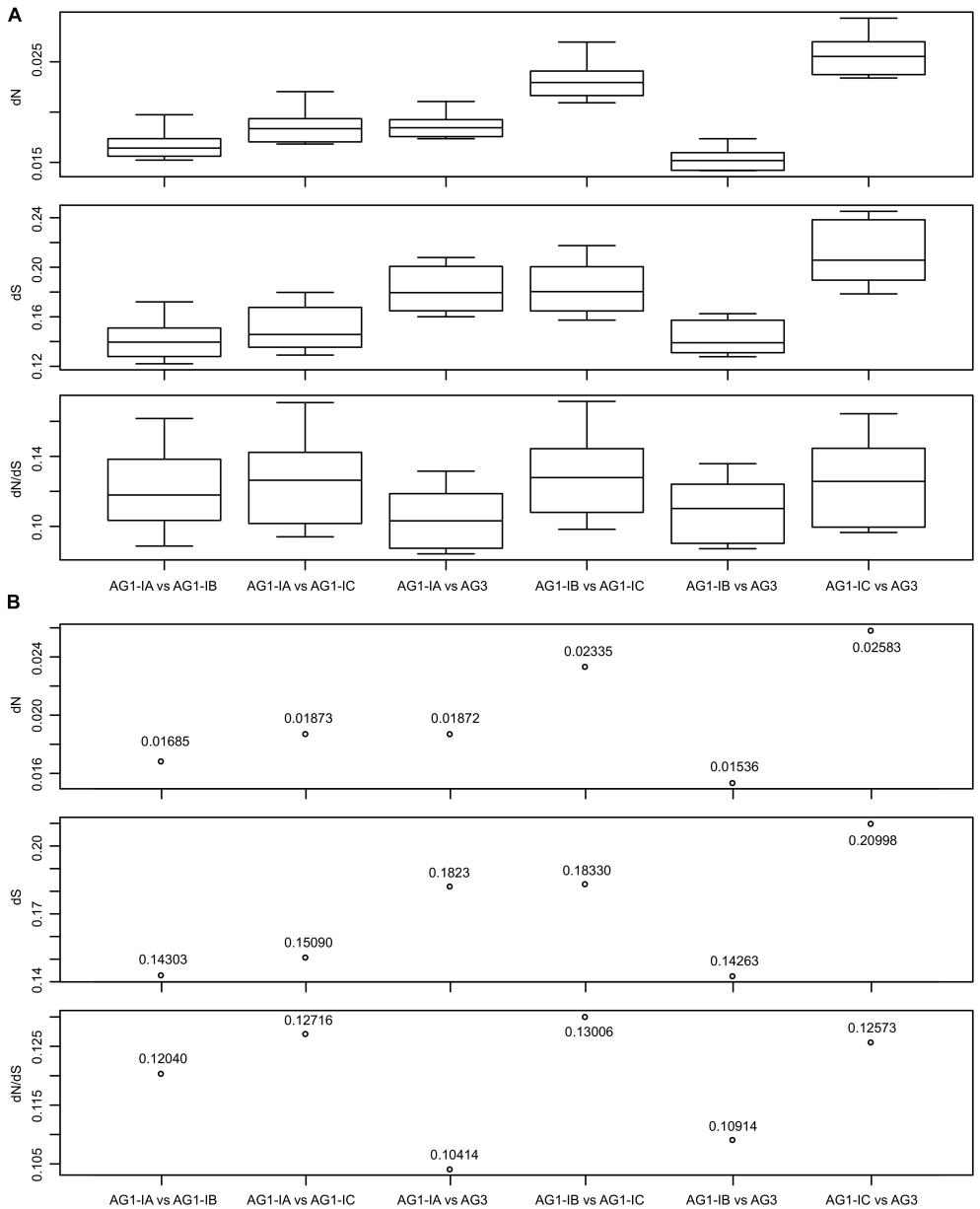


FIGURE 3 | The dN, dS, and dN/dS values for each pair of *Rhizoctonia solani* mitogenomes. **(A)** The box plots displayed the values estimated by 10 methods implemented in the KaKs_Calculator. The analysis was done based on concatenated sequences of 15 typical protein-coding genes. **(B)** The mean of the estimated values shown in **(A)**.

of the *rps3* gene showed that the contiguous sites with positively selected signals were in the helix and coil regions (**Supplementary Figure 2A**). For these sites, we marked them in the predicted protein structure (with RoseTTAFold confidence of 0.46) as well (**Supplementary Figure 2B**).

Expression Patterns of Genes in *R. solani* Mitogenomes During Host Infection

From RNA-seq for fungi-host interactions (i.e., interactions between AG1-IA and rice, soybeans, corn; AG1-IB and lettuce; AG3 and potatoes; Zheng et al., 2013; Xia et al., 2017; Zrenner et al., 2020), we analyzed the expression patterns of mitochondrial genes, which may suggest their roles during host infection. Based on gene expression FPKM values, 51 and 27 expressed genes from AG1-IA and AG3 mitogenomes, respectively, were all clustered into four clusters (**Figures 5A,B** and **Supplementary Tables 12,13**). Not all of each functional group of genes cytochrome c oxidase subunit, ATP synthase subunit, NADH dehydrogenase subunit, LAGLIDADG endonuclease, and GIY-YIG endonuclease were clustered into the same groups. For example, *cox1*, *cox2*,

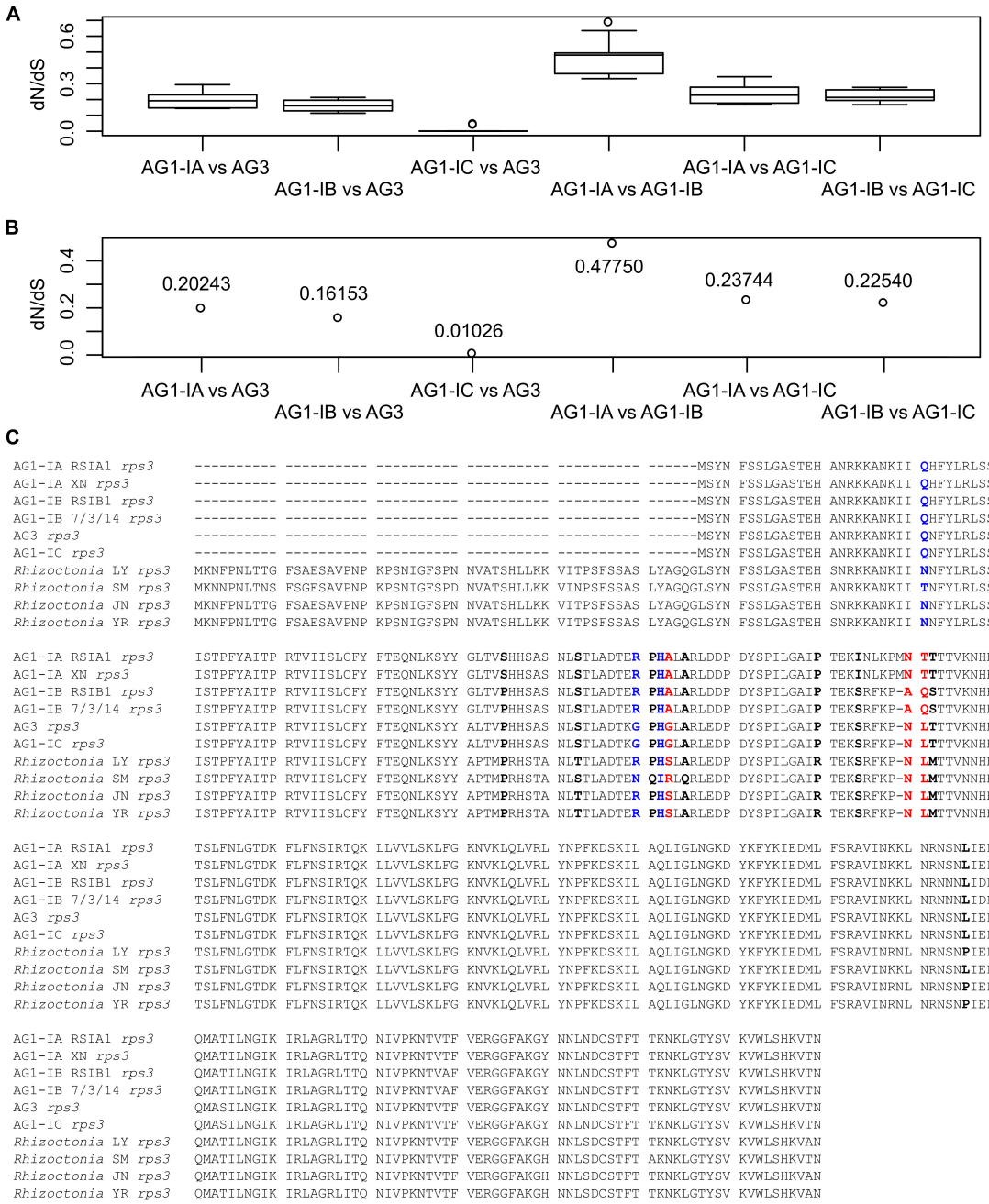


FIGURE 4 | Detection of positively selected sites in *rps3*. **(A)** The dN/dS values for each pair of *rps3* genes from mitogenomes. The dN/dS values were calculated by 10 methods implemented in the KaKs_Calculator. **(B)** The mean of the estimated values shown in **(A)**. **(C)** Display of positively selected sites in *rps3* genes. The 14 positively selected sites were detected by the CODEML program in PAML, including three sites (in blue) with $P > 95\%$ and three sites (in red) with $P > 99\%$.

and *cox3* from AG1-IA were clustered into three groups. The gene clusters showed different expression patterns after infecting hosts.

We further explored gene expression during rice infection, which displayed varied gene expression peaks (**Figure 5C**). The peaks for *cox1*, *cox2*, and *cox3* from AG1-IA mitogenome were at 24 (i.e., 24 h), 10, and 18-h, respectively, although they all had functional cytochrome c oxidase subunits. Similarly, the peaks for *atp6* and *atp9* were at 48 and 32-h, respectively, and *atp8* was not expressed during rice infection. Meanwhile, LAGLIDADG endonucleases and GIY-YIG endonucleases represented expression peaks after host infection. For example, *mag2* (a LAGLIDADG endonuclease) displayed an expression pattern similar to that of *nad6*, with the peak at 32-h; mag6 (a GIY-YIG endonuclease) showed a peak at 18-h. Similarly, peaks for different genes from AG3

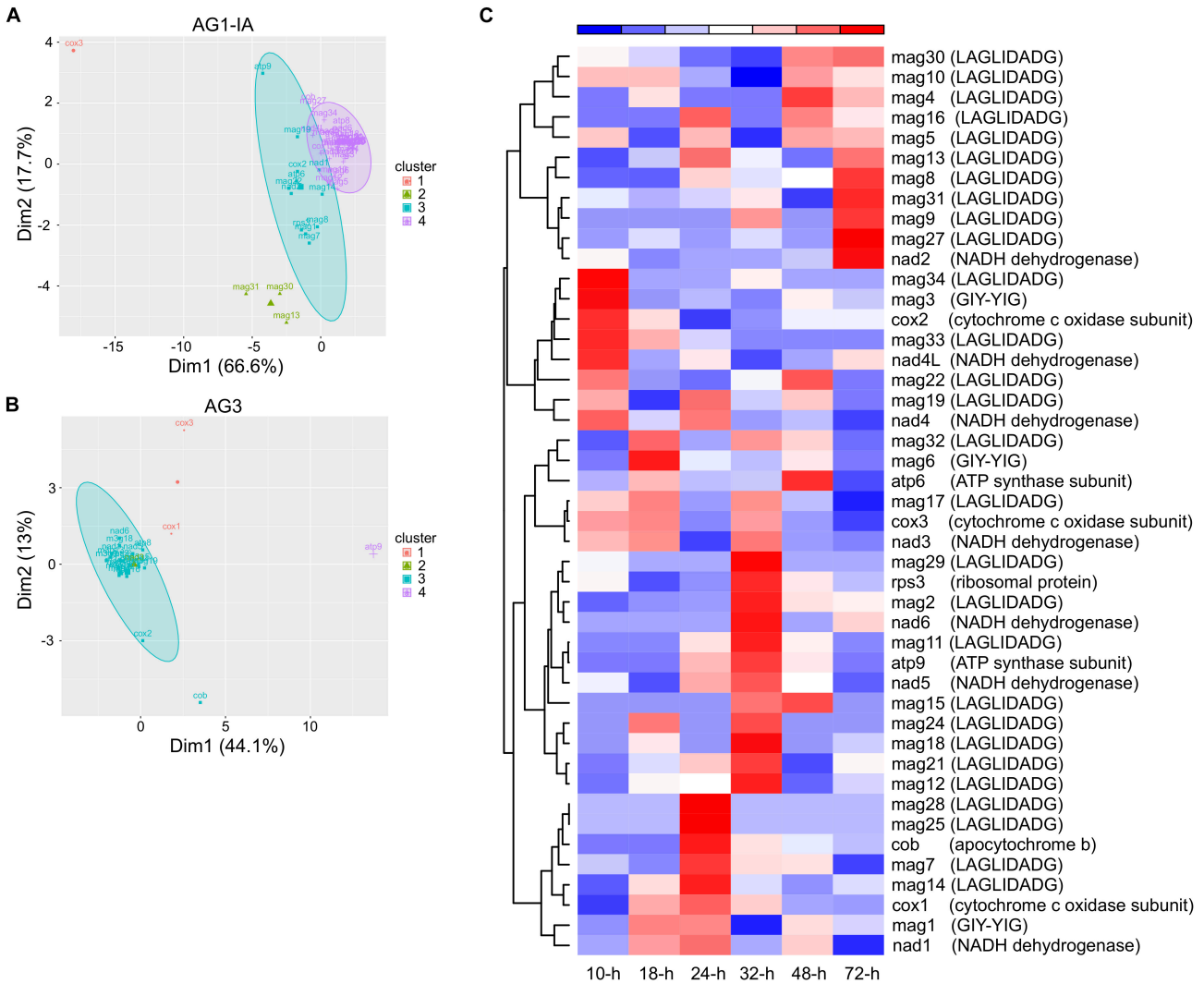


FIGURE 5 | The expression patterns of genes from *Rhizoctonia solani* mitogenomes. **(A,B)** The distribution of gene clusters from AG1-IA and AG3. **(C)** The heatmap of AG1-IA genes after infecting rice at 10, 18, 24, 32, 48, and 72 h.

mitogenomes during potato infection were found as well (**Supplementary Figure 3**).

Correlations Between Expressed Genes in *R. solani* Mitogenomes During Host Infection

For the expressed genes (including 15 typical proteins, LAGLIDADG endonucleases, and GIY-YIG endonucleases), we calculated the Pearson correlation coefficient to measure their expression similarity. With the Pearson correlation coefficient value cutoff of 0.6, we identified 76 and 8 positive correlations between each pair of genes in AG1-IA and AG3, respectively (**Supplementary Tables 14,15**). Among these genes, *cox1*, *cob*, *nad1*, *nad2*, *nad6*, *rps3*, and LAGLIDADG endonucleases in RSMOG16 orthologous group (11 genes in AG1-IA: *mag10*, *mag12*, *mag13*, *mag14*, *mag20*, *mag23*, *mag26*, *mag28*, *mag30*, *mag32*, *mag36*; three genes in AG3: *m3g4*, *m3g10*, and *m3g18*; **Supplementary Table 4**) were found in both AG1-IA and AG3. To view their relationships clearly, we chose the network to display these correlated pair of genes. The different topologies for these correlations were shown in AG1-IA and AG3 (**Figure 6**), with only two conserved correlations (i.e., *nad2-rps3* and *nad1-mag28*/*m3g18*). In the neighboring genes *nad2* and *rps3* there were some repeat sequences; however, these repeat sequences from AG1-IA and AG3 mitogenomes were not similar sequences detected by BLAST alignment. Because the correlation coefficient was inferred from gene expression during varied host infection (i.e., rice infection for AG1-IA and potato infection for AG3; **Supplementary Tables 12,13**), there were different network topologies for expressed genes (**Figure 6**), which showed their different expression profiles in AG1-IA and AG3 mitogenomes, possibly indicating the different roles for these

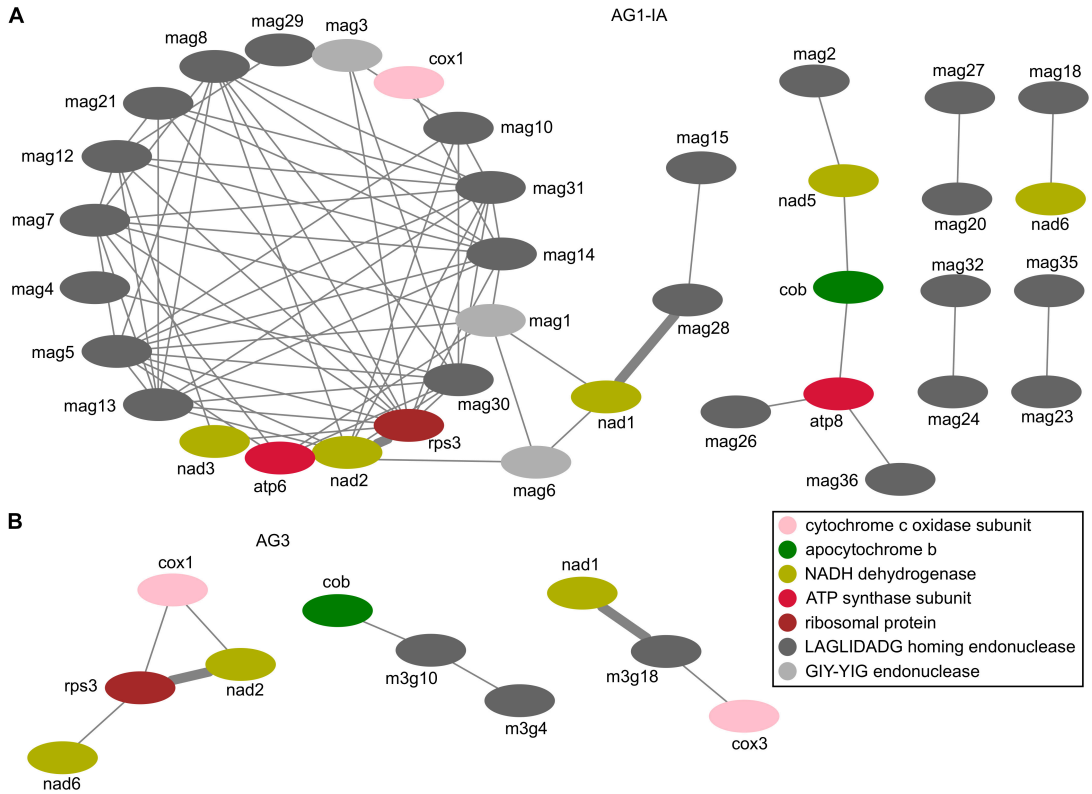


FIGURE 6 | Correlation coefficients between each pair of genes encoded in mitogenomes. **(A)** The correlations for genes in AG1-IA. **(B)** The correlations for genes in AG3.

genes after AG1-IA and AG3 mitogenomes separated from the same ancestor.

DISCUSSION

It has been proposed that mitochondria evolved from free-living bacteria via symbiosis within a eukaryotic host cell (Margulis, 1970). With that in mind, we compared examples of GC content variations in the bacterial kingdom with those we have observed in some of our fungal mitogenomes. In bacteria, the high genomic GC content is proposed to be associated with high rates of DNA damage and environmental factors (Wu et al., 2012; Weissman et al., 2019), and it is suggested to be maintained in some species by mutation pressure (Hildebrand et al., 2010). Considering both GC content and mitogenomic sizes, the Agaricomycotina clade is of interest. A comparison of mitogenomes between *Rhizoctonia* and other non-pathogens in Agaricomycotina showed that *Rhizoctonia* had a higher GC content, and a comparison of mitogenomes between *Rhizoctonia* and other plant pathogens that were separated early in Basidiomycota showed that *Rhizoctonia* had more repeat sequences. The results may indicate the divergent evolution of Basidiomycota mitogenomes. The high GC content in mitogenomes of *Rhizoctonia* that have broader host ranges may have evolved under pathogenic environmental pressure.

The comparative analysis of five complete *R. solani* mitogenomes showed the natural existence of varied mitogenomic characteristics in sizes, endonuclease genes (both LAGLIDADG and GIY-YIG endonucleases), and repeat sequences. The repeat sequences and endonucleases are the major contributors to the size variations. At least 12∼36 endonucleases were encoded in these mitogenomes (**Table 1**), and there were 73,546 bp (31.18%) repeat sequences in the AG3 mitogenome, more than those in AG1-IA (**Supplementary Tables 6,10**). In fungal mitogenomes, multiple repeat sequences are the main cause of size expansion in mitogenomes (Losada et al., 2014; Li et al., 2015). The genes encoding endonucleases are considered mobile genetic elements that invaded introns and intergenic sequences, and they have been found to play an important role in causing mitogenome size variation (Kolesnikova et al., 2019).

LAGLIDADG and GIY-YIG endonucleases have been found in fungal mitogenomes belonging to orders in all fungal phyla (Belfort et al., 2002; Megarioti and Kouvelis, 2020). These endonucleases possess special conserved amino acid motifs and are encoded in the intron regions of fungal mitogenomes (Stoddard, 2014). The LAGLIDADG endonuclease has the ability to recognize 18–22 bp target sequences (Belfort and Roberts, 1997; Chevalier et al., 2005). These endonucleases may originate from free-standing open reading frames, and endonucleases and their intron hosts may have co-evolved through recombination

and horizontal gene transfer (Megarioti and Kouvelis, 2020). Yeast endonucleases have been found to drive recombination of protein-coding genes (Wu and Hao, 2019). Currently, only five complete mitogenomes in *R. solani* have been reported. With the increasing release of genomic data, the evolution of endonucleases in *R. solani* will be explored in the future studies.

Meanwhile, the expression peaks of endonucleases during host infection were identified (**Figure 5C**), such as the high expression of *mag3* (a GIY-YIG endonuclease, located within the intergenic region between *atp6* and *rns*) and *mag33* (a LAGLIDADG endonuclease, located within the intergenic region between *nad4* and *rps3*) at 10-h after rice infection. The expression patterns of *mag31* (a LAGLIDADG endonuclease) and *rps3* containing the intron host of *mag31* were different, i.e., with expression peaks at 72 and 32-h, respectively, indicating that endonuclease and its inserted gene were expressed independently. These expression peaks may suggest the significant roles of endonucleases during host infection and independent roles for invasive endonucleases/introns and *rps3* genes. As fungal mitochondria acting as organelles to provide energy for cell functions, their encoded genes displayed varied expression peaks after infection, indicating significant cooperation among these genes.

The analysis of interspersed repeat sequences in the AG3 mitogenome suggested that the stable secondary structures exhibited by repeats may comprise catalytic RNA elements (Losada et al., 2014). None of the repeat sequences were shared between AG3 and AG1-IB 7/3/14 or between AG3 and other fungal mitogenomes in Basidiomycota, suggesting the unique evolutionary phenomenon of repeat acquisition in *R. solani* (Losada et al., 2014). The mitochondrial repeat sequences had been considered as putative elements for recombination or regulation (Ghikas et al., 2006). For both complete AG1-IB mitogenomes (AG1-IB RSIB1 and 7/3/14 in **Figure 1A**), their repeat sequence contents vary from each other, and these differences resulted in the size variation between AG1-IB and AG1-IB 7/3/14 mitogenomes (**Supplementary Tables 7,8**). Meanwhile, in the AG1-IA mitogenome, the repeat sequences may affect the expression of genes because similar repeat sequences nearby the each pair of genes (*rps3-nad2* and *nad1-mag28*) with positive correlations in expression were found. However, the influence of repeat sequence on gene expression in mitogenomes is required to further evaluate.

Non-synonymous and synonymous substitution rates (dN and dS) were different for each pair of *R. solani* mitogenomes (**Figure 3**), which may affect the phylogenetic placement of AG1-IC in the phylogeny (**Figure 2A**) because it was far from AG1-IB branches. The conflict between mitochondrial (**Figure 2A** and **Supplementary Figure 4**) and nuclear DNA (data not shown) phylogenies was identified for AG1-IC lineage. In the phylogenetic tree of nuclear genomes, AG1-IC was most closely related to AG1-IB, and they formed a clade parallel with AG1-IA. In our previous RNA-seq analysis, we found that among AG1 strains, AG1-IB and AG1-IC had the most and least frequent polymorphisms, respectively (Yamamoto et al., 2019), which was consistent with our mitogenomic analysis, i.e., the comparison of sequences between AG1-IC and other strains with high substitution rates (**Figure 3**).

Positive selection signals in fungal mitochondrial *rps3* genes have been reported previously (Lin et al., 2015, 2017; Kang et al., 2017; Wang et al., 2020; Zhang et al., 2020; Huang et al., 2021; Wu et al., 2021). Together with *rps3*, genes encoding ribosomal subunits with positive AT and GC skewness are identified in the mitogenomes of brown rot fungal pathogens (Yildiz and Ozkilinc, 2021). In our results, we detected several sites in *R. solani rps3* genes representing positively selected signals. These sites may be the hot spot region in the *R. solani* mitogenomes and they may contribute to host adaptation.

Mitochondrial DNA has been popularly used to design markers for study of genetic diversity (Galtier et al., 2009), such as the study in medicinal fungus *Cordyceps militaris* (Zhang et al., 2017). However, to our knowledge, the used of DNA markers to investigate intraspecific genetic diversity of *Rhizoctonia* sp. are mainly designed from nuclear genomes (Das et al., 2020). With the increase in publication of *Rhizoctonia* mitogenomes from different AGs, the design of mitochondrial DNA markers for identification of pathogens will become possible. Meanwhile, our mitochondrial phylogeny including AG1-IA, AG1-IB, AG1-IC, AG3, AG4, and other *Rhizoctonia* strains that adapt to different hosts will acting as a phylogenetic marker to investigate host adaptation between AGs.

Additionally, the expression of mitogenome encoded genes may offer clues to understand host adaptation for *R. solani* strains in the future studies. Although the 15 typical protein-coding genes were highly conserved in the strains, their expression in AG1-IA and AG3 during rice and potato infection were quite different (**Figures 5**, **6** and **Supplementary Figure 3**). The AG1-IA has many plant hosts, including rice, corn, soybeans, barley, potatoes, and cabbage, while AG3 hosts are potatoes and tobacco (Yang and Li, 2012). The host infection process requires energy provided by mitochondria. To adapt to different host infection, gene regulation in mitochondria may be very complex. The different correlation coefficient maps in AG1-IA and AG3 showed the more complex relationships between genes in AG1-IA (**Figure 6**). Even for AG1-IA strains, the *atp8* gene from rice isolated strains was not expressed during rice infection, while the *atp8* gene from soybeans or corn isolated strains was expressed during rice infection (**Supplementary Table 12**). These gene repertoires may be difficult to explain currently, but the strain-specific phenomena of gene expression patterns were very interesting.

Gene expression is a fundamental life process, which is essential for fungal growth, metabolism, virulence, and response to environments. The comparison of expression patterns between RNA-seq and RT-qPCR analyses (**Supplementary Figure 5**) suggested the complex expression and regulation for genes, although similar patterns were found for several genes (such as *cob*, *rps3*, *mag28*, and *mag4*). Those highly expressed genes in rice infection (**Figure 5**), such as the *cox2* with an expression peak at 10-h, may play a significant role at the beginning of AG1-IA pathogenesis and may act as candidate targets for disease control. A comparison of amino acids between AG1-IA *cox2* and human *cox2* (i.e., *MT-CO2*) showed highly conserved sequences, with *E*-value of 2e-72 and identity of 46%. The sequence mutations in human *MT-CO2* have been reported to be related to serious

diseases (Rahman et al., 1999; Heidari et al., 2020), suggesting that there may be also some potential key pathogenic factors in the *R. solani* mitogenome. The CRISPR gene-editing technology could facilitate genetic alterations in fungal genomes and enable study of gene function (Liu et al., 2015; Muñoz et al., 2019), in relation to changes in fungal growth, morphology, and virulence. Gene editing may also accelerate our understanding of the role of mitochondrial genes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the NCBI under the following accession numbers: MW995474.1 (https://www.ncbi.nlm.nih.gov/nuccore/MW995474.1/), MW995475.1 (https://www.ncbi.nlm.nih.gov/nuccore/MW995475.1/), and MW995476.1 (https://www.ncbi.nlm.nih.gov/nuccore/MW995476.1).

AUTHOR CONTRIBUTIONS

RL, YX, YL, and DZ contributed equally to this work. AZ designed the study. YX, DZ, and YL conducted the experiments. RL, YX, DZ, XX, XN, and AZ analyzed the mitogenomic data. RL, YX, DZ, YL, LJ, and XW analyzed the transcriptomic data. RL and AZ submitted the data to NCBI. RL, YX, YL, and AZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Project of Science and Technology Department of Sichuan Province (No. 2019YFN0010), the National Natural Science Foundation (No. 31400130), and the National 973 Project (No. 2014CB160304).

ACKNOWLEDGMENTS

The authors wish to thank Erxun Zhou for providing the national standard strains of *R. solani* and lab members for their helpful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.707281/full#supplementary-material

REFERENCES

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754

Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B., and Lambowitz, A. (2002). "Mobile introns: pathways and proteins" in *Mobile DNA II*. eds N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (Washington: ASM Press). 761–783. doi: 10.1128/9781555817954.ch31

Belfort, M., and Roberts, R. J. (1997). Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* 25, 3379–3388. doi: 10.1093/nar/25.17.3379

Buchan, D. W. A., and Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res*. 47, W402–W407. doi: 10.1093/nar/gkz297

Bullerwell, C. E., and Lang, B. F. (2005). Fungal evolution: the case of the vanishing mitochondrion. *Curr. Opin. Microbiol.* 8:362. doi: 10.1016/j.mib.2005.06.009

Calderone, R., Li, D., and Traven, A. (2015). System-level impact of mitochondria on fungal virulence: to metabolism and beyond. *FEMS Yeast Res.* 15:fov027. doi: 10.1093/femsyr/fov027

Carling, D. E. (1996). "Grouping in *Rhizoctonia solani* by hyphal anastomosis reaction," in *Rhizoctonia* Species: taxonomy, Molecular Biology, Ecology, Pathology and Disease Control, eds B. Sneh, S. Jabaji-Hare, S. Neate, and G. Dijst (Dordrecht: Kluwer Academic Publishers), 35–47.

Chak, S. T. C., Baeza, J. A., and Barden, P. (2021). Eusociality shapes convergent patterns of molecular evolution across mitochondrial genomes of snapping shrimps. *Mol. Biol. Evol.* 38, 1372–1383. doi: 10.1093/molbev/msaa297

Chang, A. L., and Doering, T. L. (2018). Maintenance of mitochondrial morphology in *Cryptococcus neoformans* is critical for stress resistance and virulence. *Mbio* 9:6. doi: 10.1128/mBio.01375-18

Chatre, L., and Ricchetti, M. (2014). Are mitochondria the Achilles' heel of the kingdom fungi? *Curr. Opin. Microbiol.* 20, 49–54. doi: 10.1016/j.mib.2014.05.001

Chevalier, B., Monnat, R., and Stoddard, B. (2005). "The LAGLIDADG homing endonuclease family" in *Homing Endonucleases and Inteins*. eds M. Belfort, V. Derbyshire, B. Stoddard, and D. Wood (Heidelberg: Springer-Verlag). 33–45. doi: 10.1007/3-540-29474-0_3

Ciampi, M. B., Meyer, M. C., Costa, M. J., Zala, M., McDonald, B. A., and Ceresini, P. C. (2008). Genetic structure of populations of *Rhizoctonia solani* anastomosis group-1 IA from soybean in Brazil. *Phytopathology* 98, 932–941. doi: 10.1094/PHYTO-98-8-0932

Cubeta, M., Thomas, E., Dean, R., Jabaji, S., Neate, S., Tavantzis, S., et al. (2014). Draft genome sequence of the plant-pathogenic soil fungus *Rhizoctonia solani* anastomosis group 3 strain Rhs1AP. *Genome Announc.* 2, e01072–14. doi: 10.1128/genomeA.01072-14

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

Das, S., Plyler-Harveson, T., Santra, D. K., Maharjan, B., Nielson, K. A., and Harveson, R. M. (2020). A longitudinal study on morpho-genetic diversity of pathogenic *Rhizoctonia solani* from sugar beet and dry beans of western Nebraska. *BMC Microbiol.* 20:354. doi: 10.1186/s12866-020-02026-9

Duvenage, L., Walker, L. A., Bojarczuk, A., Johnston, S. A., MacCallum, D. M., Munro, C. A., et al. (2019). Inhibition of classical and alternative modes of respiration in *Candida albicans* leads to cell wall remodeling and increased macrophage recognition. *Mbio* 10:1. doi: 10.1128/mBio.02535-18

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32, 1792–1797. doi: 10.1093/nar/gkh340

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* 33, 243–246. doi: 10.1038/nbt.3172

Fu, R., Zhou, M., Yin, C., Zheng, A., and Li, P. (2014). Karyotype analysis and research of infection process of *Rhizoctonia solani* AG-1 IC. *J. Pure Appl. Microbiol.* 8, 1209–1216.

Galtier, N., Nabholz, B., Glémin, S., and Hurst, G. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18, 4541–4550. doi: 10.1111/j.1365-294X.2009.04380.x

Gonzalez Garcia, V., Portal Onco, M. A., and Rubio Susan, V. (2006). Review. Biology and systematics of the form genus *Rhizoctonia*. *Span. J. Agric. Res.* 4, 55–79. doi: 10.5424/sjar/2006041-178

Ghikas, D. V., Kouvelis, V. N., and Typas, M. A. (2006). The complete mitochondrial genome of the entomopathogenic fungus *Metarhizium anisopliae* var. *anisopliae*: gene order and trn gene clusters reveal a common evolutionary course for all Sordariomycetes, while intergenic regions show variation. *Arch. Microbiol.* 185, 393–401. doi: 10.1007/s00203-006-0104-x

Grahl, N., Dinamarco, T. M., Willger, S. D., Goldman, G. H., and Cramer, R. A. (2012). *Aspergillus fumigatus* mitochondrial electron transport chain mediates oxidative stress homeostasis, hypoxia responses and fungal pathogenesis. *Mol. Microbiol.* 84, 383–399. doi: 10.1111/j.1365-2958.2012.08034.x

Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* 4:a011403. doi: 10.1101/cshperspect.a011403

Hane, J. K., Anderson, J. P., Williams, A. H., Sperschneider, J., and Singh, K. B. (2014). Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS Genet.* 10:e1004281. doi: 10.1371/journal.pgen.1004281

Heidari, M. M., Mirfakhradini, F. S., Tayefi, F., Ghorbani, S., Khatami, M., and Hadadzadeh, M. (2020). Novel point mutations in mitochondrial *MT-CO2* gene may be risk factors for coronary artery disease. *Appl. Biochem. Biotechnol.* 191, 1326–1339. doi: 10.1007/s12010-020-03275-0

Hildebrand, F., Meyer, A., and Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet.* 6:e1001107. doi: 10.1371/journal.pgen.1001107

Huang, W., Feng, H., Tu, W., Xiong, C., Jin, X., Li, P., et al. (2021). Comparative mitogenomic analysis reveals dynamics of intron within and between *Tricholoma* species and phylogeny of Basidiomycota. *Front. Genet.* 12:534871. doi: 10.3389/fgene.2021.534871

Ingavale, S. S., Chang, Y. C., Lee, H., McClelland, C. M., Leong, M. L., and Kwon-Chung, K. J. (2008). Importance of mitochondria in survival of *Cryptococcus neoformans* under low oxygen conditions and tolerance to cobalt chloride. *PLoS Pathog.* 4:e1000155. doi: 10.1371/journal.ppat.1000155

Kang, X., Hu, L., Shen, P., Li, R., and Liu, D. (2017). SMRT sequencing revealed mitogenome characteristics and mitogenome-wide DNA modification pattern in *Ophiocordyceps sinensis*. *Front. Microbiol.* 8:1422. doi: 10.3389/fmicb.2017.01422

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kolesnikova, A. I., Putintseva, Y. A., Simonov, E. P., Biriukov, V. V., Oreshkova, N. V., Pavlov, I. N., et al. (2019). Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related *Armillaria* species. *BMC Genom.* 20:351. doi: 10.1186/s12864-019-5732-z

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 k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Lee, D. Y., Jeon, J., Kim, K. T., Cheong, K., Song, H., Choi, G., et al. (2021). Comparative genome analyses of four rice-infecting *Rhizoctonia solani* isolates reveal extensive enrichment of homogalacturonan modification genes. *BMC Genomics* 22:242. doi: 10.1186/s12864-021-07549-7

Li, C., Guo, Z., Zhou, S., Han, Q., Zhang, M., Peng, Y., et al. (2021a). Evolutionary and genomic comparisons of hybrid uninucleate and nonhybrid *Rhizoctonia* fungi. *Commun. Biol.* 4:201. doi: 10.1038/s42003-021-01724-y

Li, C., Zhou, S., Zhang, M., Guo, Z., and Chen, X. (2021b). Comparison of mitochondrial genomes from multi-, Bi-, and uninucleate *Rhizoctonia*. *Mitochondrial DNA B Resour.* 6, 472–474. doi: 10.1080/23802359.2021.1872430

Li, Y., Hu, X. D., Yang, R. H., Hsiang, T., Wang, K., Liang, D. Q., et al. (2015). Complete mitochondrial genome of the medicinal fungus *Ophiocordyceps sinensis*. *Sci. Rep.* 5:13892. doi: 10.1038/srep13892

Lin, R., Liu, C., Shen, B., Bai, M., Ling, J., Chen, G., et al. (2015). Analysis of the complete mitochondrial genome of *Pochonia chlamydosporia* suggests a close relationship to the invertebrate-pathogenic fungi in Hypocreales. *BMC Microbiol.* 15:5. doi: 10.1186/s12866-015-0341-8

Lin, R., Cheng, X., and Xie, B. (2017). "Comparative analysis of *Pochonia chlamydosporia* mitogenome reveals dynamic mitochondrial evolution of the nematophagous fungi in Hypocreales" in *Perspectives in Sustainable Nematode Management through Pochonia Chlamydosporia Applications for Root and Rhizosphere Health. Sustainability in Plant and Crop Protection*. eds R. Manzanilla-López and L. Lopez-Llorca (Germany: Springer). 183–195. doi: 10.1007/978-3-319-59224-4_9

Liu, R., Chen, L., Jiang, Y., Zhou, Z., and Zou, G. (2015). Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. *Cell Discov.* 1:15007. doi: 10.1038/celldisc.2015.7

Losada, L., Pakala, S. B., Fedorova, N. D., Joardar, V., Shabalina, S. A., Hostetler, J., et al. (2014). Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiol. Lett.* 352, 165–173. doi: 10.1111/1574-6968.12387

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 0955–0964. doi: 10.1093/nar/25.5.955

Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven: Yale University Press.

Megarioti, A. H., and Kouvelis, V. N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.* 12, 1337–1354. doi: 10.1093/gbe/evaa126

Misawa, T., and Aoki, M. (2017). First report of *Rhizoctonia solani* AG-1 IC causing head rot of cabbage in Japan. *N. Dis. Rep.* 36:12. doi: 10.5197/j.2044-0588.2017.036.012

Molla, K. A., Karmakar, S., Molla, J., Bajaj, P., Varshney, R. K., Datta, S. K., et al. (2020). Understanding sheath blight resistance in rice: the road behind and the road ahead. *Plant Biotechnol. J.* 18, 895–915. doi: 10.1111/pbi.13312

Monteiro-Vitorello, C. B., Bell, J. A., Fulbright, D. W., and Bertrand, H. (1995). A cytoplasmically transmissible hypovirulence phenotype associated with mitochondrial DNA mutations in the chestnut blight fungus *Cryphonectria parasitica*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5935–5939. doi: 10.1073/pnas.92.13.5935

Muñoz, I. V., Sarrocco, S., Malfatti, L., Baroncelli, R., and Vannacci, G. (2019). CRISPR-Cas for fungal genome editing: a new tool for the management of plant diseases. *Front Plant Sci.* 10:135. doi: 10.3389/fpls.2019.00135

Nadarajah, K., Mat Razali, N., Cheah, B. H., Sahruna, N. S., Ismail, I., and Tathode, M. (2017). Draft genome sequence of *Rhizoctonia solani* anastomosis group 1 subgroup 1A strain 1802/KB isolated from rice. *Genome Announc.* 5, e01188–17. doi: 10.1128/genomeA.01188-17

Nielsen, R., and Yang, Z. (2003). Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* 20, 1231–1239. doi: 10.1093/molbev/msg147

Ogoshi, A. (1987). Ecology and pathogenicity of anastomosis and intraspecific groups of *Rhizoctonia solani* Kuhn. *Annu. Rev. Phytopathol.* 25, 125–143. doi: 10.1146/annurev.py.25.090187.001013

Pannecoucque, J., and Höfte, M. (2009). Interactions between cauliflower and *Rhizoctonia* anastomosis groups with different levels of aggressiveness. *BMC Plant Biol.* 9:95. doi: 10.1186/1471-2229-9-95

Patil, V., Girimalla, V., Sagar, V., Bhardwaj, V., and Chakrabarti, S. (2018). Draft genome sequencing of *Rhizoctonia solani* anastomosis group 3 (AG3-PT) causing stem canker and black scurf of potato. *Am. J. Potato Res.* 95, 87–91. doi: 10.1007/s12230-017-9606-0

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21, 351–358. doi: 10.1093/bioinformatics/bti1018

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Opens external link in new window. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rahman, S., Taanman, J. W., Cooper, J. M., Nelson, I., Hargreaves, I., Meunier, B., et al. (1999). A missense mutation of cytochrome oxidase subunit II causes defective assembly and myopathy. *Am. J. Hum. Genet.* 65, 1030–1039. doi: 10.1086/302590

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Sneh, B., Burpee, L., and Ogoshi, A. (1991). *Identification of Rhizoctonia Species*. USA: APS Press.

Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* 6:31. doi: 10.1186/1471-2105-6-31

Stoddard, B. L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* 5:7. doi: 10.1186/1759-8753-5-7

Sun, N., Parrish, R. S., Calderone, R. A., and Fonzi, W. A. (2019). Unique, diverged, and conserved mitochondrial functions influencing *Candida albicans* respiration. *Mbio* 10, e00300–19. doi: 10.1128/mBio.00300-19

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Verwaaijen, B., Wibberg, D., Krober, M., Winkler, A., Zrenner, R., Bednarz, H., et al. (2017). The *Rhizoctonia solani* AG1-IB (isolate 7/3/14) transcriptome during interaction with the host plant lettuce (*Lactuca sativa* L.). *PLoS One* 12:e0177278. doi: 10.1371/journal.pone.0177278

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, X., Jia, L., Wang, M., Yang, H., Chen, M., Li, X., et al. (2020). The complete mitochondrial genome of medicinal fungus *Taiwanofungus camphoratus* reveals gene rearrangements and intron dynamics of Polyporales. *Sci. Rep.* 10:6500. doi: 10.1038/s41598-020-73461-x

Weissman, J. L., Fagan, W. F., and Johnson, P. L. F. (2019). Linking high GC content to the repair of double strand breaks in prokaryotic genomes. *PLoS Genet.* 15:e1008493. doi: 10.1371/journal.pgen.1008493

Wibberg, D., Jelonek, L., Rupp, O., Hennig, M., Eikmeyer, F., Goesmann, A., et al. (2013). Establishment and interpretation of the genome sequence of the hytopathogenic fungus *Rhizoctonia solani* AG1-IB isolate 7/3/14. *J. Biotechnol.* 167, 142–155. doi: 10.1016/j.jbiotec.2012.12.010

Wibberg, D., Rupp, O., Jelonek, L., Kröber, M., Verwaaijen, B., Blom, J., et al. (2015). Improved genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB 7/3/14 as established by deep mate-pair sequencing on the MiSeq (Illumina) system. *J. Biotechnol.* 203, 19–21. doi: 10.1016/j.jbiotec.2015.03.005

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

Wu, H., Zhang, Z., Hu, S., and Yu, J. (2012). On the molecular mechanism of GC content variation among eubacterial genomes. *Biol. Direct.* 7:2. doi: 10.1186/1745-6150-7-2

Wu, P., Bao, Z., Tu, W., Li, L., Xiong, C., Jin, X., et al. (2021). The mitogenomes of two saprophytic Boletales species (*Coniophora*) reveals intron dynamics and accumulation of plasmid-derived and non-conserved genes. *Comput. Struct. Biotechnol. J.* 19, 401–414. doi: 10.1016/j.csbj.2020.12.041

Xia, Y., Fei, B., He, J., Zhou, M., Zhang, D., Pan, L., et al. (2017). Transcriptome analysis reveals the host selection fitness mechanisms of the *Rhizoctonia solani* AG1IA pathogen. *Sci. Rep.* 7:10120. doi: 10.1038/s41598-017-10804-1

Yamamoto, N., Wang, Y., Lin, R., Liang, Y., Liu, Y., Zhu, J., et al. (2019). Integrative transcriptome analysis discloses the molecular basis of a heterogeneous fungal phytopathogen complex, *Rhizoctonia solani* AG-1 subgroups. *Sci. Rep.* 9:19626. doi: 10.1038/s41598-019-55734-2

Yang, G., and Li, C. (2012). "General description of *Rhizoctonia* species complex" in *Plant Pathology*. ed. C. Cumagun (Rijeka: InTech). 41–52.

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

Yang, Z., and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917. doi: 10.1093/oxfordjournals.molbev.a004148

Yildiz, G., and Ozkilinc, H. (2021). Pan-mitogenomics approach discovers diversity and dynamism in the prominent brown rot fungal pathogens. *Front. Microbiol.* 12:647989. doi: 10.3389/fmicb.2021.647989

Zhang, S., Hao, A. J., Zhao, Y. X., Zhang, X. Y., and Zhang, Y. J. (2017). Comparative mitochondrial genomics toward exploring molecular markers in the medicinal fungus *Cordyceps militaris*. *Sci. Rep.* 7:40219. doi: 10.1038/srep40219

Zhang, Y., Yang, G., Fang, M., Deng, C., Zhang, K. Q., Yu, Z., et al. (2020). Comparative analyses of mitochondrial genomes provide evolutionary insights into nematode-trapping fungi. *Front. Microbiol.* 11:617. doi: 10.3389/fmicb.2020.00617

Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteom. Bioinform.* 4, 259–263. doi: 10.1016/S1672-0229(07)60007-2

Zhang, Z., Xia, X., Du, Q., Xia, L., Ma, X., Li, Q., et al. (2021). Genome sequence of *Rhizoctonia solani* anastomosis group 4 strain Rhs4ca, a wide spread pathomycete in field crops. *Mol. Plant Microbe. Interact.* 19:MPMI12200362A.doi: 10.1094/MPMI-12-20-0362-A

Zheng, A., Lin, R., Zhang, D., Qin, P., Xu, L., Ai, P., et al. (2013). The evolution and pathogenic mechanisms of the rice sheath blight pathogen. *Nat. Commun.* 4:1424. doi: 10.1038/ncomms2427

Zrenner, R., Genzel, F., Verwaaijen, B., Wibberg, D., and Grosch, R. (2020). Necrotrophic lifestyle of *Rhizoctonia solani* AG3-PT during interaction with its host plant potato as revealed by transcriptome analysis. *Sci. Rep.* 10:12574. doi: 10.1038/s41598-020-68728-2

Conflict of Interest: The 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.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 December 2021
doi: 10.3389/fmicb.2021.787283

Global Characterization of Fungal Mitogenomes: New Insights on Genomic Diversity and Dynamism of Coding Genes and Accessory Elements

Paula L. C. Fonseca[1,2†‡], Ruth B. De-Paula[3†‡], Daniel S. Araújo[4†‡], Luiz Marcelo Ribeiro Tomé[5†], Thairine Mendes-Pereira[5†], Wenderson Felipe Costa Rodrigues[6†], Luiz-Eduardo Del-Bem[6,7†], Eric R. G. R. Aguiar[2*†] and Aristóteles Góes-Neto[5,6*†]*

[1] Department of Genetics, Ecology and Evolution, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] Department of Biological Science (DCB), Center of Biotechnology and Genetics (CBG), Universidade Estadual de Santa Cruz (UESC), Ilhéus, Brazil, [3] Graduate School of Biomedical Sciences, Baylor College of Medicine, Houston, TX, United States, [4] Program in Bioinformatics, Loyola University Chicago, Chicago, IL, United States, [5] Molecular and Computational Biology of Fungi Laboratory, Department of Microbiology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [6] Program of Bioinformatics, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [7] Department of Botany, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Fungi comprise a great diversity of species with distinct ecological functions and lifestyles. Similar to other eukaryotes, fungi rely on interactions with prokaryotes and one of the most important symbiotic events was the acquisition of mitochondria. Mitochondria are organelles found in eukaryotic cells whose main function is to generate energy through aerobic respiration. Mitogenomes (mtDNAs) are double-stranded circular or linear DNA from mitochondria that may contain core genes and accessory elements that can be replicated, transcribed, and independently translated from the nuclear genome. Despite their importance, investigative studies on the diversity of fungal mitogenomes are scarce. Herein, we have evaluated 788 curated fungal mitogenomes available at NCBI database to assess discrepancies and similarities among them and to better understand the mechanisms involved in fungal mtDNAs variability. From a total of 12 fungal phyla, four do not have any representative with available mitogenomes, which highlights the underrepresentation of some groups in the current available data. We selected representative and non-redundant mitogenomes based on the threshold of 90% similarity, eliminating 81 mtDNAs. Comparative analyses revealed considerable size variability of mtDNAs with a difference of up to 260 kb in length. Furthermore, variation in mitogenome length and genomic composition are generally related to the number and length of accessory elements (introns, HEGs, and uORFs). We identified an overall average of 8.0 (0–39) introns, 8.0 (0–100) HEGs, and 8.2 (0–102) uORFs per genome, with high variation among phyla. Even though

OPEN ACCESS

Edited by:
Vassili N. Kouvelis,
National and Kapodistrian University of Athens, Greece

Reviewed by:
Nicolaas A. van der Merwe,
University of Pretoria, South Africa
Baojun Wu,
Clark University, United States

***Correspondence:**
Luiz-Eduardo Del-Bem
delbem@ufmg.com
Eric R. G. R. Aguiar
ericgdp@gmail.com
Aristóteles Góes-Neto
arigoesneto@icb.ufmg.br

†ORCID:
Paula L.C. Fonseca
orcid.org/0000-0002-8103-9235
Ruth B. De-Paula
orcid.org/0000-0003-1022-2056
Daniel S. Araújo
orcid.org/0000-0001-6219-9463
Luiz Marcelo Ribeiro Tomé
orcid.org/0000-0002-1477-5558
Thairine Mendes-Pereira
orcid.org/0000-0002-3707-9790
Wenderson Felipe Costa Rodrigues
orcid.org/0000-0002-6167-0900
Luiz-Eduardo Del-Bem
orcid.org/0000-0001-8472-4476
Eric R. G. R. Aguiar
orcid.org/0000-0002-8143-5756
Aristóteles Góes-Neto
orcid.org/0000-0002-7692-6243

[‡] These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 30 September 2021
Accepted: 11 November 2021
Published: 01 December 2021

the length of the core protein-coding genes is considerably conserved, approximately 36.3% of the mitogenomes evaluated have at least one of the 14 core coding genes absent. Also, our results revealed that there is not even a single gene shared among all mitogenomes. Other unusual genes in mitogenomes were also detected in many mitogenomes, such as *dpo* and *rpo,* and displayed diverse evolutionary histories. Altogether, the results presented in this study suggest that fungal mitogenomes are diverse, contain accessory elements and are absent of a conserved gene that can be used for the taxonomic classification of the Kingdom Fungi.

Keywords: mitochondrial genome, homing endonuclease, intron, size variation, comparative genomics

INTRODUCTION

The Kingdom Fungi is one of the most diverse and globally distributed groups of eukaryotes. Fungi evolved a wide morphological, physiological, and ecological heterogeneity that allowed them to perform vital functions in both terrestrial and aquatic ecosystems (Walker and White, 2005; James et al., 2020; Li et al., 2021). The global fungal diversity is estimated between 2 to 5 million species, that are further classified into 12 phyla: Ascomycota, Basidiomycota, Entorrhizomycota, Chytridiomycota, Monoblepharidomycota, Neocallimastigomycota, Mucoromycota, Zoopagomycota, Blastocladiomycota, Aphelidiomycota, Cryptomycota/Rozellomycota, and Microsporidia (James et al., 2020; Li et al., 2021).

Estimations indicate that only ∼10% of the fungal diversity has been described, with the subkingdom Dikarya containing ∼97% of all known species, further distributed into the phyla Ascomycota, Basidiomycota, and the recently added Entorrhizomycota (James et al., 2020; Li et al., 2021). Ascomycota alone encompasses ∼70% of all Dikarya diversity (Schmidt-Dannert, 2016), suggesting that most groups of fungi are still underrepresented. Thus, integrative studies to identify and describe new species, as well as those applying large phylogenies and sequencing of genomes (both nuclear and mitochondrial), are still needed to better understand the evolution, biology, and phylogeny of this highly important group of organisms.

Eukaryotes and prokaryotes are in constant interaction, and the most ancient symbiosis we know about led to the emergence of mitochondria (Carvalho et al., 2015). Mitochondria are bi-membranous cytoplasmic organelles responsible for a vast range of functions. Their central function is the production of energy, via oxidative phosphorylation and adenosine triphosphate (ATP) synthesis (Kulik et al., 2021). Mitochondria also produce metabolic precursors of macromolecules, such as proteins and lipids, and metabolic by-products, such as ammonia and reactive oxygen species (Spinelli and Haigis, 2018). In addition, these organelles are involved in apoptosis, homeostasis maintenance, and stress response (Grahl et al., 2012; Verma et al., 2018; Zardoya, 2020). It is proposed that mitochondria originated from the endosymbiosis of an ancestral free-living Alphaproteobacteria, which was permanently integrated into the host cell (Carvalho et al., 2015; Muñoz-Gómez et al., 2017). Mitochondria contain their own genome (mitochondrial DNA – mtDNA or mitogenome), with independent replication and inheritance. Throughout evolutionary history, most of the mitochondrial early genes have been transferred to the nuclear genome. As a result, most mitochondrial proteins are encoded by the nucleus, translated by ribosomes in the cytosol, and subsequently transferred into the mitochondria. Therefore, extant mtDNA contains mainly protein-coding genes that play central roles in oxidative phosphorylation (Verma et al., 2018). The mitogenome and the nuclear genome are in constant communication, during which the mtDNA is responsible for the stability of the nuclear genome and proper cell functioning. The mitogenome also depends on the nuclear genome to maintain its functionality, integrity, and stability (Kaniak-Golik and Skoneczna, 2015; Kulik et al., 2021).

The architecture of mitogenomes can vary among eukaryotes. In fungi, mitogenomes can be linear or circular. Circular mitogenomes are by far the most common, while linear mitogenomes are mainly described in yeasts, such as *Saccharomyces cerevisiae* (Ascomycota) (Williamson, 2002; Formaggioni et al., 2021). Typical fungal mitogenomes such as those belonging to Ascomycota and Basidiomycota (Dikarya Subkingdom) usually display a core genome with 14 conserved protein-coding genes (*atp6*, *8*, and *9; cob*; *cox1– 3;* and *nad1– 6, 4L*), the ribosomal genes *rnl* and *rns*, and from 20 and 31 tRNA genes (Zardoya, 2020; Kulik et al., 2021). The genomes belonging to subkingdom Dikarya also have the ribosomal protein gene *rps3* that is lacking in some taxa, such as the genus *Candida* (Freel et al., 2015; Formaggioni et al., 2021). Mitogenome length is also quite variable among fungal taxa, ranging from 12 to 272 kbp in *Rozella allomycis* and *Morchella crassipes*, respectively (James et al., 2013; Liu et al., 2020). This plasticity in mtDNA is related to factors like the length and number of introns, intergenic regions, repetitive DNA, plasmid insertions, segment duplications, Open Reading Frames (ORFs) without defined function (uORF), and homing endonuclease genes (HEG) (Fonseca et al., 2020; Zardoya, 2020; Araújo et al., 2021; Kulik et al., 2021).

The presence of accessory DNA elements, such as introns, HEGs, and uORFs are directly correlated to the modulation of mtDNA (Fonseca et al., 2020; Araújo et al., 2021). In fungal mitogenomes, group I and group II introns can be identified. The classification of these introns is based on sequence conservation, secondary structure configuration, and helix shape (Hausner et al., 2014). Group I introns are the most frequent and can encode HEGs families like LAGLIDADG

and GIY-YIG, which are related to transposition of sequences in different regions of the mitogenome (Hafez and Hausner, 2012; Kulik et al., 2021). Group II introns differ from type I in structure, sequence, and splicing mechanism. The distribution of introns among different species is considered irregular, and the same introns and their associated HEGs can be identified in mitogenomes that are considered evolutionarily distant (Hausner, 2003; Lang et al., 2007; Hausner et al., 2014; Zardoya, 2020).

In spite of the diversity of size and genes, mitogenomes can be a key tool to investigate phylogenetic relationships by exhibiting a low recombination rate, diverse patterns of inheritance, and by displaying a mutation rate different from the nuclear genome (Rubinoff and Holland, 2005; Mendoza et al., 2020). Additionally, the high number of copies inside the cytoplasm of the cell facilitates the amplification and sequencing of partial or complete mitogenomes (Kulik et al., 2021). Nonetheless, the sequencing of mitogenomes from a larger number of fungal species is necessary to assess the presence of a universal marker to understand the evolution of major fungal lineages (Kulik et al., 2015; Smith, 2015; Avin et al., 2017).

Considering the increasing number of fungal mitogenomes available in genomic databases in the last few years due to exponential growth of high-throughput sequencing (HTS), we observed a discrepancy regarding the diversity of fungal taxa represented. As fungi display a great diversity of species with a broad range of ecological functions, we selected the Kingdom Fungi as a model to better understand the macroevolutionary patterns underlying mitogenome diversity. We characterized the fungal mitogenomes according to genome size, content, and sequence homology to better understand the mechanisms involved in the observed sequence diversity among fungal mtDNAs. We also correlated the genetic code with mitogenome topology, genome size, and gene composition of core and accessory elements. Some species of the 12 fungal phyla are overrepresented; meanwhile, at least four phyla have no representative in genomic databases. In addition, we found a high variation in mitogenome length in fungi. This variation may be due to the variable abundance of accessory (introns, HEGs and uORFs) and protein-coding regions. Our analyses showed that no gene is universally conserved in fungal mitogenomes, suggesting that mitochondrial gene content can vary widely without disrupting organelle function. The data explored herein provides a significant addition to our understanding of the diversity and evolution of fungal mitogenomes.

MATERIALS AND METHODS

Bibliographic Survey for Fungal Mitogenomes Studies

A literature review was carried out to obtain information on the number and objective of studies published with fungal mitogenomes until December 2020. The search was performed with the keywords "((fungi[Title/Abstract]) OR (fungal[Title/Abstract])) AND ((mitochondria[Title/Abstract]) OR (mitochondria[Title/Abstract])) AND genomic" in the PubMed database. The abstract of each study was evaluated and classified by year of publication and by topic: characterization of mitogenomes (description of a new mitogenome, comparative analyses) or biotechnological (study of mitogenomes in industrial or antifungal processes). The **Supplementary Table 1** presents all data from the resulting articles.

Inclusion Criteria of Mitogenomes

A total of 788 mitogenomes were available for download in the NCBI database[1] on December 4th, 2020. In our analyses, we selected representative and non-redundant mitogenomes based on the threshold of 90% identity using the CD-HIT software (Li and Godzik, 2006). This threshold was defined after a comparison of the results obtained with 70, 80, and 90–99% similarity among the genomes (**Supplementary Table 2**). Two mitogenomes (NC_003061.1 and NC_003060.1), classified in NCBI as complete but with less than 2 kbp in size, were removed. Another 22 mitogenomes of poor quality containing many gaps (more than 20% of N bases) were also removed from our analyses. The selected mitogenomes ($n = 685$) were used in all subsequent analyses. Identification of fungal species, taxonomic classification, and accession numbers for the mitogenomes are provided in **Supplementary Table 3**.

Mitochondrial Genome Annotation

We verified each species' page in NCBI Taxonomy Browser[2] to assess the proper mitochondrial genetic code for protein-coding gene annotation. Mitogenomes were annotated using Mfannot[3] to standardize the process and to allow a fairer comparative analysis. Mfannot program was selected for its greater specificity in the annotation of fungal mitochondrial genes that has already been used in other studies of mitogenome characterization (Chen et al., 2019; Fonseca et al., 2020; Araújo et al., 2021). Mitogenomes that did not display genes considered essential (*rns*, *rnl*, *rps3*, *atp*, *cox* and *nad* families, and *cob*) were manually curated. A gene was considered absent when it did not show nucleotide similarity above 50% when compared to all mitochondrial genes deposited in the NCBI database[4]. All annotations are available in **Supplementary File 1**.

Annotation of uORFs

With the annotation files generated by Mfannot, we extracted the sequences annotated as "ORFs" and translated them into amino acids using the corresponding mitochondrial genetic code for each species. These protein sequences were then submitted to the Batch CD-Search interface of NCBI's Conserved Domain Database[5] to identify known protein domains in the unidentified open reading frames.

Sequences that did not exhibit similarity with known domains were considered as uORFs and clustered into a single file.

[1]https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/

[2]https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

[3]http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl

[4]https://www.ncbi.nlm.nih.gov/

[5]https://www.ncbi.nlm.nih.gov/cdd/

Sequences that presented identity above 80% with HEGs, DNA-polymerase (*dpo*), and RNA-polymerase (*rpo*), were clustered into different files for further analyses. Furthermore, we estimated the coding potential of all annotated ORFs (with and without known protein domains) using the CPC2 software (Kang et al., 2017). The ORF sequences classified as *rpo* or *dpo* were aligned with MAFFT (Katoh et al., 2019). The selection of the best nucleotide substitution model was performed using MEGA X software (Kumar et al., 2018) based on AIC criteria (Akaike, 1973). The best model selected for both datasets was GTRCAT. A maximum likelihood phylogeny with 1,000 bootstrap replicates was generated using MEGA X. The output files were used to plot the trees using the *ggtree* package v.3.0.2 (Yu, 2020) in R (R Core Team, 2021).

Comparative Mitogenome Analysis

The comparative analysis consisted of evaluating the difference in length of mitogenomes of each phylum, the number of genes, the presence of core genes, the presence and length of introns, HEGs, uORFs, GC content, and similarity amongst the mitogenomes. We have developed in-house scripts to parse software outputs and calculate genome summary statistics to facilitate tabulating the data.

Initially, we wrote a Python script that uses Biopython's modules to compute the length and GC content of every mitogenome analyzed in this study. Using the output files from Mfannot annotation, we designed a Bash script to generate tabular files containing the annotation data ("feature name," "start coordinate," "end coordinate," "feature size," and "feature strand"). According to "feature name," output files (tabular and fasta format) were generated for multiple categories: ORFs, introns, HEGs, core genes (including the *rps3* ribosomal protein), tRNA genes and other unusual genes found in mitogenomes (*dpo* and *rpo*). The number of features per file was counted to get the total number of features per category. Another Python algorithm that removes overlaps of feature coordinates and calculates final genic lengths was written to properly calculate the length of the intergenic spaces in fungal mitogenomes. These numbers were subtracted from the total fungal mitochondrial genome length, resulting in the total length of intergenic regions. The data obtained from the aforementioned scripts were organized and plotted using the *ggplot2* package v.3.3.5 (Wickham, 2016) in R (R Core Team, 2021).

Mitogenome Similarity Network Analyses

For the similarity network analyses, we first used BLASTn (Johnson et al., 2008) to assess similarity among mitogenomes. Then, we built two different networks: the first one consisted of only overrepresented mitogenomes (species that present more than three mitogenomes after the similarity filter made by CD-HIT), and the second one with all non-redundant mitogenomes used in this study, including the mitogenomes from the first network. In cases in which mitogenome pairs aligned in more than one segment, that is, the same pair aligned multiple times in different regions, we only considered the similarity index of the longest continuous alignment. Subsequently, we built the networks using the *ggraph* package v.2.0.5 (Pedersen and RStudio, 2021) in R (R Core Team, 2021).

RESULTS AND DISCUSSION

Fungal Mitogenomes Overview

Fungal mitogenomes can be very informative to describe new species, identify genes related to fungicides, and to help understand the process of energy production and general cell physiology (Kulik et al., 2021). Although the genetic study of fungal mitochondria is of great relevance, literature about this topic is far from abundant (481 papers found in the last 40 years in our bibliographic search) (**Supplementary Table 1**). Most published articles are related only to the description of a new mitogenome or comparative mitogenomes analysis from a particular group of fungi, while a few others propose mitochondrial genes as molecular markers or the importance of mitogenomes for specific fungal species. Comparatively, according to Smith (2015), until 2015 more than 1,000 articles had been published for other kingdoms addressing issues of mitogenome description, use of mitochondrial information as a molecular marker, and origin and ancestry of these organisms (Smith, 2015). Considering the PubMed database, using the keyword Mitogenome or Mitochondrial genome, 5,302 articles are listed, of which 3,267 were published from 2016 to date.

Until December 4th, 2020, there were 788 fungal mitogenomes available in the NCBI database. The vast majority were deposited in the last 8 years (**Figure 1A**). The increase of mitogenomes in the last years can be explained by the emergence and improvements of HTS. However, the vast majority of the 147,933 identified fungal species (Kirk, 2020) still do not have their mitogenomes sequenced (Cheek et al., 2020). The number of mitogenomes available in the database does not reflect the fungal diversity, suggesting that further sequencing studies should be carried out to investigate the genomic architecture of the kingdom. Compared to Metazoa by the year 2014, there were already more than 4,000 complete mitogenome entries available in NCBI, representing more than 90% of all mitogenomes available in the database.

Different biases in fungal genomic analyses have been reported, with the most important regarding the substantially higher number of studies on Ascomycota species compared to species from other phyla. It is estimated that at least 63% of fungal species belong to Ascomycota (Kirk, 2020) although an even higher proportion of the fungal genomics literature is dedicated to this phylum. This may be due to the fact that this phylum comprises the largest number of species with high economic importance and diverse ecological functions, and species considered as model organisms, providing new information regarding the biology and evolution of the Kingdom Fungi (Robbertse et al., 2006; Schoch et al., 2009; Wallen and Perlin, 2018). Our search reflected this bias for fungal mitochondrial DNA sequences: from the total of 788 NCBI entries, Ascomycota has twice as many deposits as other phyla, with many species displaying more than one sequenced mitogenome. Moreover, some phyla are poorly sampled, such

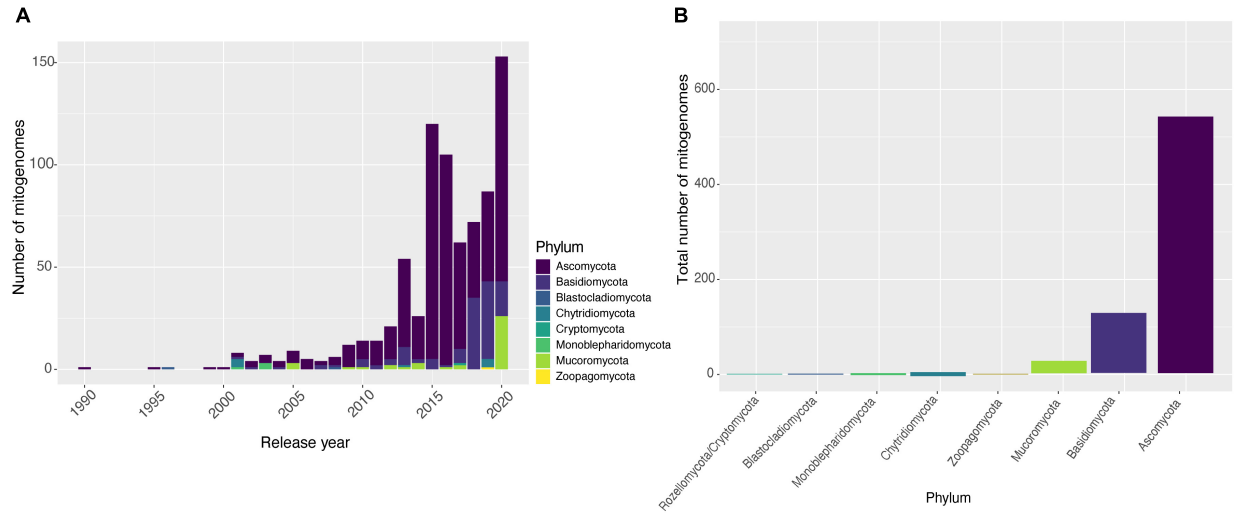


FIGURE 1 | Availability of fungal mitogenomes in NCBI databases. **(A)** Number of mitogenomes published by year and classified by phyla, from 1990 to 2020. **(B)** Number of mitogenomes available for each phylum of the Kingdom Fungi.

as Blastocladiomycota (two mitogenomes) and Zoopagomycota (three mitogenomes) (**Figure 1B**). This low sampling of the other phyla limits our knowledge on the composition and evolution of fungal mitogenomes. Nevertheless, it should still be considered that some fungal phyla have a much smaller number of species described when compared to Ascomycota and Basidiomycota, as is the case of Cryptomycota/Rozellomycota (Lazarus and James, 2015; Letcher and Powell, 2018; Li et al., 2021).

In this study, we aimed to evaluate the differences and similarities among the mitogenomes of the Kingdom Fungi using the CD-HIT software to select a dataset without the presence of highly similar sequences. At a 90% similarity, the first inflection point was observed, in which there was a reduction of the dataset by 10.28% (removal of 81 access numbers). Despite the removal of these sequences, the dataset remained with almost 90% of all mitogenomes, suggesting that the genomic content is quite divergent, which led us to investigate the conservation of protein-coding genes and the presence of accessory elements (introns, HEGs and uORFs) in most of the mitogenomes.

Plasticity of Fungal Mitogenomes

Mitogenome Length Variation and Gene Conservation

Many studies have shown that fungal mitogenomes have great structural plasticity (Fonseca et al., 2020; Araújo et al., 2021; Kulik et al., 2021). In this study, the length difference among mitogenomes ranges from 12 (*R. allomycis* Chytridiomycota – NC_003061) to 272 kbp (*Ophiocordyceps camponoti* Ascomycota – CM022976). Since the two mitogenomes are from different and distant phyla, we assessed whether mitogenome length would be related to phyla classification (**Figure 2A**). In general, Ascomycota (63,888 ± 35,298 bp) and Basidiomycota (77,394 ± 49,928 bp) mitogenomes have quite variable lengths. However, Cryptomycota (25,401 bp) and Mucoromycota (61,936 ± 14,895 bp) have smaller and more homogeneous sizes. This result may be related to the availability of data, since more than 90% are composed of mitogenomes from Ascomycota and Basidiomycota, while Mucoromycota and Cryptomycota represent 3.38% and 0.13% of mitogenomes, respectively.

As the length of mitogenomes is variable and is related to variations in both protein-coding and accessory element regions (Paquin et al., 1997; Bullerwell, 2003; Al-Reedy et al., 2012; Araújo et al., 2021), we re-annotated all mitogenomes to facilitate the comparison between the features present in each genome. After the annotation, we estimated the presence/absence of mitochondrial genes that are considered essential in fungal mitogenomes. Several previous studies have already demonstrated the presence of *atp*, *cox*, *cob*, *nad*, *rns,* and *rnl* in fungal mitogenomes (Song et al., 2020; Kulik et al., 2021). According to our analyses, although most of the mitogenomes have all the previously expected core genes, there was not a single gene shared in all evaluated fungal species. **Figure 2B** displays the frequency of each gene per phylum. In this figure, the phylum Cryptomycota is not shown, as it has only one representative (*Paramicrosporidium saccamoebae* – CM008827) containing all genes except for *nad6*. Genes of *atp*, *cox,* and *cob* families are the most frequent among fungal mitogenomes. Homologs of *nad*, *rns* and *rnl* are the least found within Ascomycota. It is worth noting that 149 mitogenomes are from the genus *Saccharomyces*, which does not have *nad* homologs (Ruan et al., 2017). In addition, some *S. cerevisiae* mitogenomes also have no homologs of *atp* and *cob* (**Figure 3**). Many studies have already reported that the coding content of the genus *Saccharomyces* is well conserved (Freel et al., 2015; Ruan et al., 2017). No member of *Saccharomycetaceae* were shown to have *nad* homologs, and studies suggested that *nad* homologs were lost after the divergence of this family (Dujon, 2010).

As a controversy of the relationship between size and presence of core genes, the mitogenome of the yeast *Malassezia furfur* (CP046241) is 43 kb long but does not show any of the

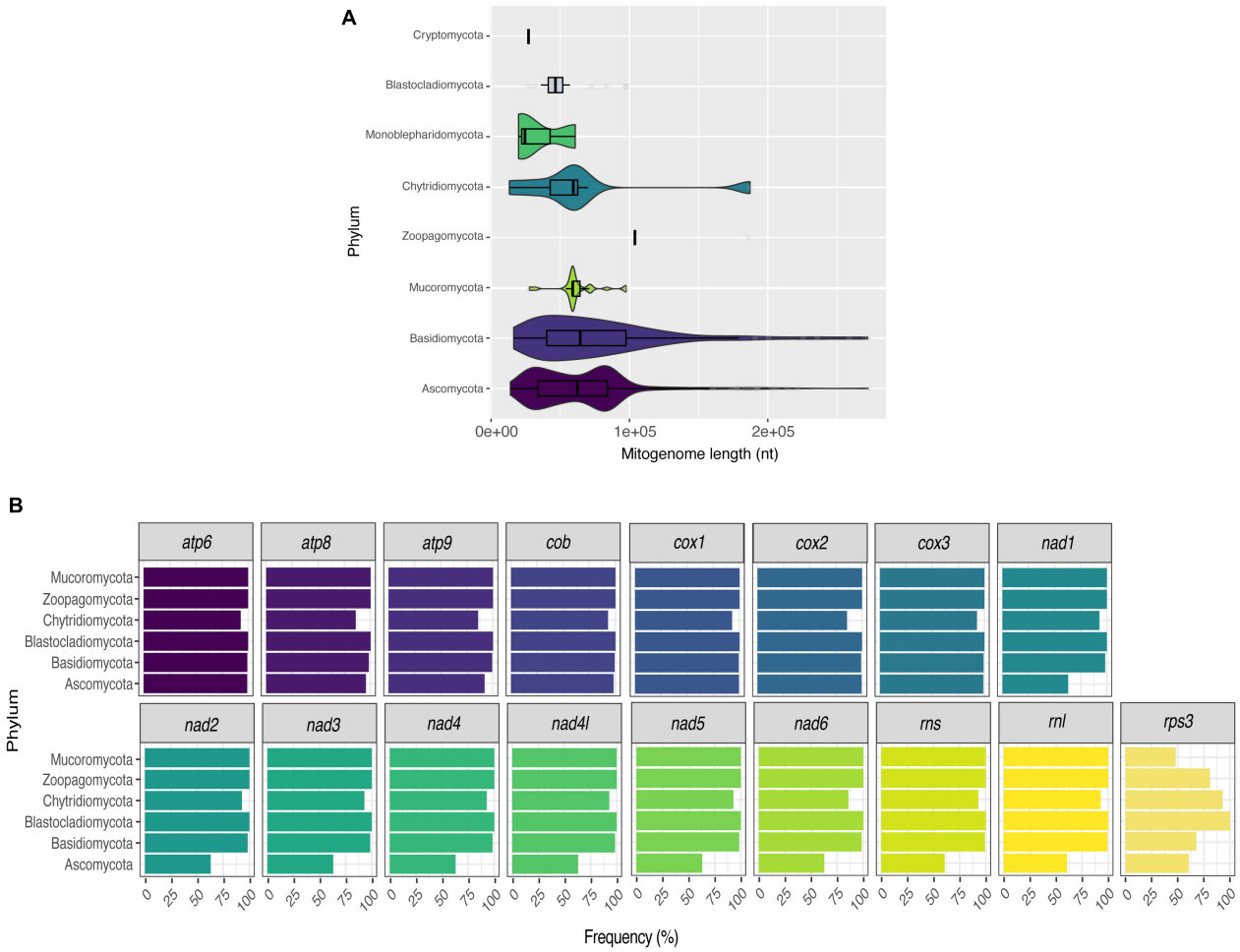


FIGURE 2 | Size and composition of fungal mitogenomes. Size **(A)** and frequency of core genes **(B)** of fungal mitogenomes by phyla.

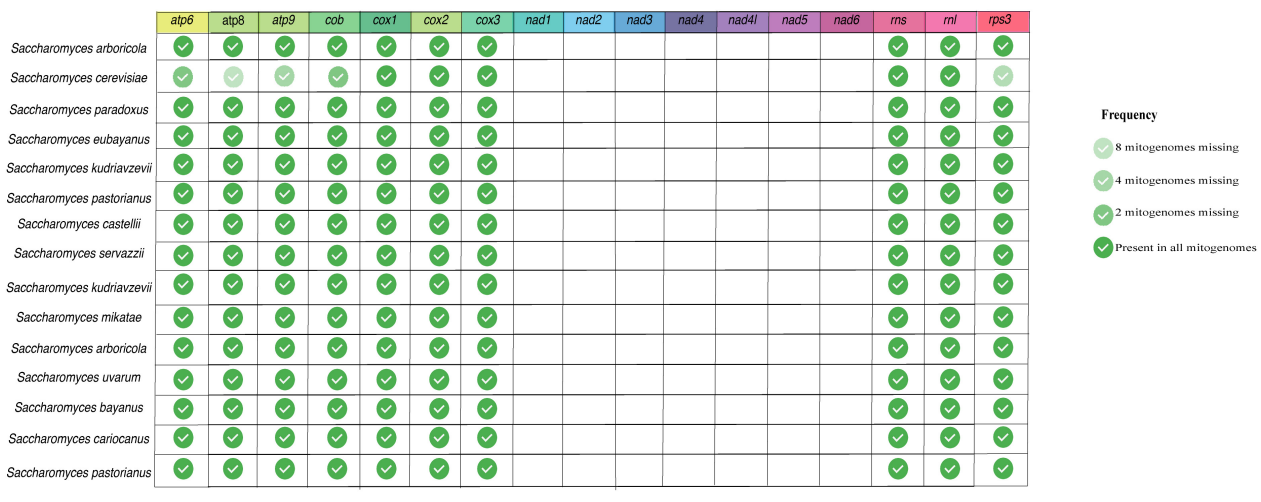


FIGURE 3 | Presence of core genes in *Saccharomyces* mitogenomes. Fourteen protein-coding and two ribosomal genes were evaluated in the 160 curated *Saccharomyces* mitogenomes.

core genes. This mitogenome has 47 ORFs, of which 22 have unknown protein domains, and the others have domains from the superfamilies haloacid dehydrogenase (HAD), Glutenin high molecular weight, among others (**Supplementary File 2**). The species *M. furfur* is a basidiomycotan yeast of medical importance as it can cause skin and blood infections (Theelen et al., 2018), and the mitogenome used in this study was previously used for taxonomic classification of the species (Sankaranarayanan et al., 2020). The genes found in this mitogenome, despite having mitochondrial-related functions, are not commonly found in mitogenomes, but in nuclear genomes (Theelen et al., 2018). Another species lacking many mitochondrial genes is *Rozella allomyces* which has the smallest mitogenome size in our dataset (12 kbp), and lacks nine core genes (*atp6*, *nad1-6,* and *rnl*).

During evolution, many mitochondrial genes migrated to the nuclear genome. In fungi, the majority of genes related to mitochondrial function are found in the nuclear genome (Bolender et al., 2008). The main theory for the mechanism involved in the escape of DNA to the cytosol and, consequently, to the nucleus states that these genetic fragments are transferred by mobile elements that use the non-homologous (NHEJ) machinery for integration (Berg and Kurland, 2000; Tsuji et al., 2012). In *S. cerevisiae*, transfer of mtDNA to the nuclear genome has already been observed when the cell had mutations in specific nuclear genes, depending on the structure of the mitogenome and availability of sugars to allow fermentation in the medium, which could lead to the high variation of phenotypes to anaerobic environments (Shafer et al., 1999; Peter et al., 2018). The absence of core genes in the species evaluated may be an indication that these genes migrated to the nuclear genome. Studies exploring the nuclear genome of these species must be carried out to confirm the presence of genes of mitochondrial origin in the nuclear genome.

Presence and Abundance of Non-coding Elements

After assessing the number and length of the following accessory elements: introns, HEGs, and uORFs, in all the fungal mitogenomes filtered by CD-HIT, we characterized the number of HEGs found per genome and grouped by phylum. Many studies have been discussing the presence and influence of non-coding elements in mitogenomes of different fungal groups (Fonseca et al., 2020; Araújo et al., 2021; Kulik et al., 2021). HEGs are classified as selfish elements that can transpose by breaking the DNA at specific sites, which consequently generates gene conversion events. These accessory elements can be found in introns or as free-standing, which ensures their dissemination and fixation in the mitogenome and may even alter the reading frame of protein-coding genes (Hausner, 2003; Barzel et al., 2011; Stoddard, 2014).

In our analysis, we identified Ascomycota species that can have up to 100 HEG elements, such as *Nemania diffusa* (NC_049077). The other phyla exhibited a more homogeneous number of HEGs, with a mean of six HEGs per mitogenome (**Figure 4A**). HEGs have not been identified in many mitogenomes, such as *Acremonium fuci* (NC_029851) and *Hyaloraphidium curvatum* (NC_003048). Currently, six families of HEGs are known, with only two families found in fungal mitogenomes, LAGLIDADG and GIY-YIG (Hausner, 2003; Megarioti and Kouvelis, 2020). We identified a total of 4,432 LAGLIDADG and 1,247 GIY-YIG across all mitogenomes. Interestingly, we found 15 H-N-H elements, showing that these HEGs occur in the mitogenome of fungi, although they are rare (Hafez and Hausner, 2012; Megarioti and Kouvelis, 2020). H-N-H are usually found as free-standing or in group I introns (IGI) in Bacteria and their phages (Hafez and Hausner, 2012). All three aforementioned types of HEGs are able to cleave double-stranded DNA at specific sites, facilitating the mobility of the elements within the genome (Arbuthnot, 2015). Since numerous HEG elements were found in most mitogenomes of fungi, we asked whether there is a correlation between the mitogenome total length, and the total length of the fraction composed by HEGs. We found a significant positive Pearson correlation ($R = 0.37$; $p < 2.2e\text{-}16$) suggesting that the abundance of HEGs length is one of the factors responsible for the observed variation in fungal mitogenome size (**Figure 4B**). However, our data indicate that they may not be the only factor, leading us to investigate whether other non-coding elements would also be responsible for the variation in size and diversity in the composition of mitogenomes.

In total, 5,516 intronic sequences were identified. Ascomycota and Basidiomycota had 3,976 (72.1%) and 1,191 (21.59%) of all introns identified, respectively. Moreover, Ascomycota mitogenomes have an average of nine introns per mitogenome while the other phyla averaged less than five (**Figure 5A**). Mitogenomes with a high number of introns were identified, such as *Pyrrhoderma noxium* (Basidiomycota – CM008263) with 39 introns, *Phycomyces blakesleeanus* (Ascomycota- NC_027730) with 38 introns, *Parasitella parasítica* (Mucoromycota – NC_024944) with 22 introns, and *Rhizophydium* sp. (Chytridiomycota – NC_003053) with 17 introns. We found that the presence and global length of introns is correlated with the size of each mitogenome ($R = 0.48$, $p < 2.2e\text{-}16$) (**Figure 5B**). Many introns contain domains of HEGs, and since these elements can transpose to other parts of the genome, we investigated whether the length of introns was related to the HEGs length (**Figure 5C**), and found a correlation $R = 0.86$, which suggests that HEGs may be partly responsible for intron size, such as proposed previously (Edgell, 2009).

We assessed the percentage of HEGs within introns as well as the opposite (**Supplementary Table 4**). We found that 43.46% of all HEGs are within intronic regions (31.54% of GIY-YIG; 33.33% of HNH; and 46.99% of LAGLIDADG), and 48.23% of the analyzed introns overlap with HEGs coordinates, and, most frequently, LAGLIDADG (9.00% overlap GIY-YIG; 0.22% overlap H-N-H; and 39.01% overlap LAGLIDADG). In fact, these results agree with the literature, which shows that 30% of IGI contain internal ORFs that might represent HEGs (Chevalier and Stoddard, 2001; Hafez and Hausner, 2012). Most probably, introns originated before HEGs as first suggested by Loizos et al. (1994), providing a "safe haven" for these endonucleases to get established, although this theory does not explain why there is still many free-standing HEGs (Loizos et al., 1994). Edgell (2009) states that both HEGs and introns are selfish, independent elements, but they often benefit from being in close genomic proximity (e.g., HEGs are established by introns, and introns

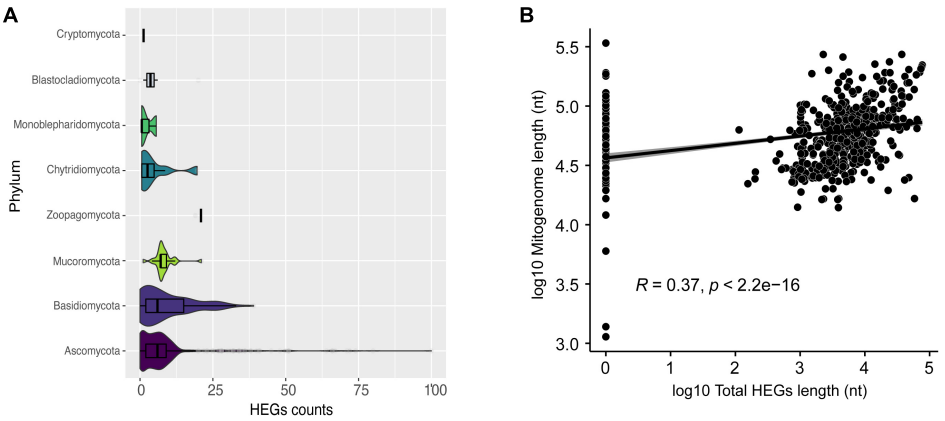


FIGURE 4 | Characterization of HEGs. **(A)** Number of HEGs found in the mitogenomes by phylum. **(B)** Correlation between mitogenome length and total HEGs length.

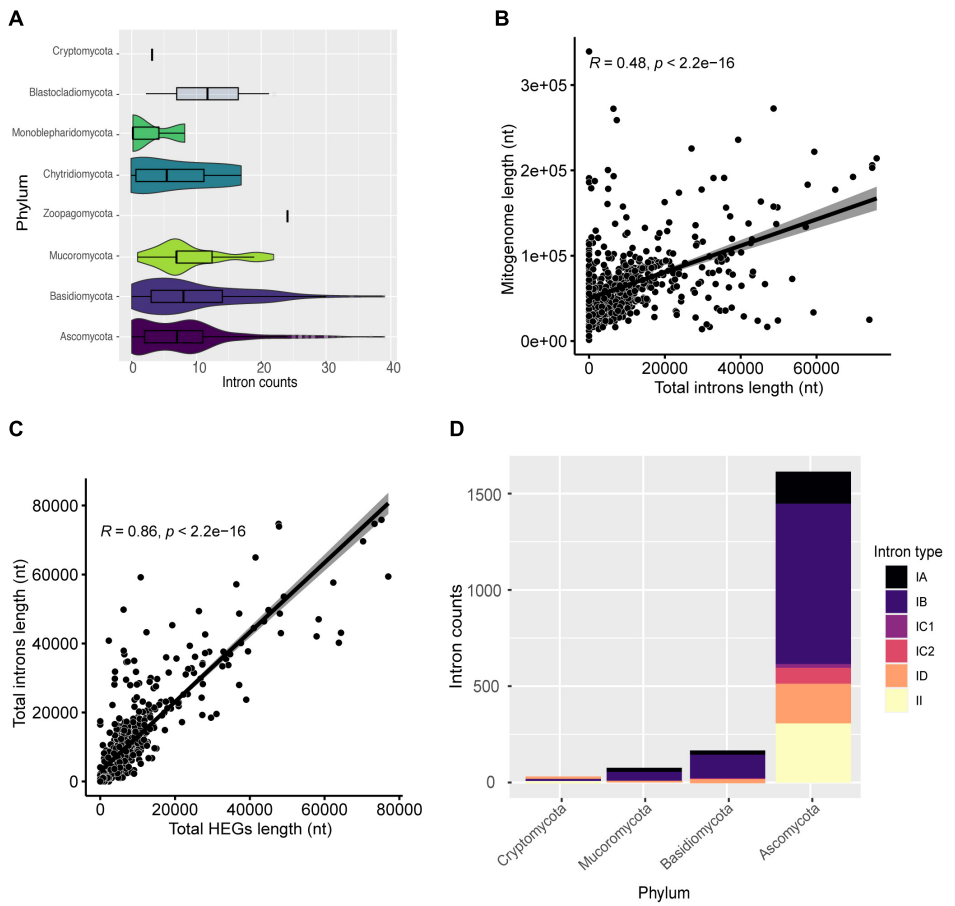


FIGURE 5 | Characterization of introns. **(A)** Number of introns found in the mitogenomes by phylum. **(B)** Correlation between mitogenome length and total introns length. **(C)** Correlation between total introns length and total HEGs length. **(D)** The most frequent types of introns in the evaluated mitogenomes classified by phylum.

can use the endonuclease properties of HEGs to spread more easily) (Edgell, 2009). The association of HEGs with introns allows their persistence in the genome, since if they inserted themselves into genes, they would be subject to negative selective pressure to eliminate the selfish elements from the genome (Stoddard, 2014).

Generally, IGI is more common in fungi while IGII is more common in plants and most of them are self-splicing

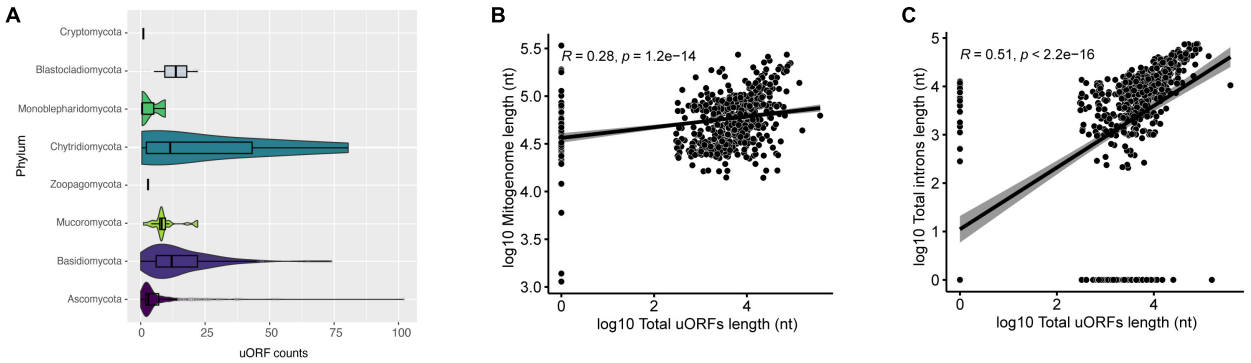


FIGURE 6 | Characterization of uORFs. **(A)** Number of uORFs found in mitogenomes in each phylum. **(B)** Correlation between mitogenome length and total uORFs length. The correlation is in log10. **(C)** Correlation between total introns length and total uORFs length. The correlation is in log10.

(Hausner et al., 2014; Fonseca et al., 2020). In this study, we performed the classification of identified introns and found six types (IA, IB, IC1, IC2, ID, and II). Introns from group IB were the most common among mitogenomes (562 or 76% of mitogenomes), while group IC1 was the least frequent (31 or 4.19% of mitogenomes). Moreover, when IGI is colonized by HEG domains, it can split a gene region due to the HEG cutting function (Lambowitz and Perlman, 1990; Belfort, 2003). Although IGII are usually infrequent, they were identified in 238 (32.21%) mitogenomes (**Figure 5D**). IGII introns have catalytic (ribozyme) and intron-encoded-protein sites, which allow their own splicing and proliferation in the genome (Lambowitz and Zimmerly, 2011). A previous study carried out by our group showed that intronic sequences can be shared among distinct species and at different genomic positions, indicating that the presence of self-splicing introns associated with maturases, such as HEGs, may be responsible for the transfer and sharing between genomes (Fonseca et al., 2020).

When we evaluated the presence of uORFs in mitogenomes, species of Ascomycota had the highest number of uORFs. For example, *Morchella importuna* (NC_012621) had 102 uORF elements and a mitogenome length of 107 kbp (**Figure 6A**). Mitogenome and uORF lengths present a weak positive linear relationship (R = 0.28, p = 1.2 e-14) (**Figure 6B**), and a moderate positive linear correlation with introns length (**Figure 6C**). These uORF elements have been identified in different fungal mitogenomes, and some may be shared among species, as recently demonstrated (Fonseca et al., 2020; Araújo et al., 2021; Kulik et al., 2021), which suggests that uORFs within mitogenomes of fungi may play a significant role in shaping mitochondrial genome length. This is in agreement with a recent investigation of mitochondrial landscape in eukaryotes, which has found a positive correlation between mitochondrial genome length and the prevalence of unassigned regions (Formaggioni et al., 2021). Nonetheless, differences in the literature suggest that there are probably inter-phyla variations. For instance, the linear correlation coefficient is greater than 0.50 in both the Agaricomycetes class and the Hypocreales order (Fonseca et al., 2020; Araújo et al., 2021), a higher value than the one obtained in this study. As aforementioned, uORFs are not homogeneously present in all phyla analyzed, and therefore they may not influence mitogenome length at the same magnitude for different groups.

Diversity of Domains in uORFs

In our attempt to identify uORFs, we found that 64.49% of them have known protein domains (**Supplementary Table 5**) and most of them were sequences with HEGs domains (69.05%). We also identified some uORFs with protein domains related to plasmid genes: *dpo* (5.60%) and *rpo* (1.05%). A total of 35.51% of the uORFs lacked any similarity with known protein domains and may represent unidentified protein-coding genes. Nevertheless, by implementing a computational technique that assess coding potential based on sequence intrinsic features, 43.86% of uORFs were predicted to be coding (**Supplementary Table 6**). Since we were working with a kingdom-wide dataset, it was out of the scope of our study to conduct genus- or species-level transcriptomic analyses to verify if those uORFs are transcribed. When possible, this method can help shed light on whether uORFs are in fact transcribed or not, as demonstrated by Torriani et al. (2014) for some *Rhynchosporium* species (Torriani et al., 2014).

Mitochondrial plasmids are widely found in Fungi and some mitochondrial genes, such as *dpo* and *rpo*, are commonly associated with plasmid integration events (Formighieri et al., 2008; Férandon et al., 2013; Formaggioni et al., 2021). In our study, we identified *dpo* homologs in at least one species of six different phyla, with Basidiomycota and Ascomycota showing the highest numbers of sequences – 227 and 116, respectively (**Supplementary Table 5**). On the other hand, *rpo* genes were only annotated in Ascomycota (46 sequences), Basidiomycota (23), and Chytridiomycota (6). In total, amongst all mitogenomes analyzed, it was possible to observe that 10.74% of Ascomycota fungi presented a *dpo* and/or *rpo* homolog, in contrast with 57.14% of Basidiomycota species. For the other phyla, the lack of representatives makes difficult to draw meaningful conclusions. For instance, most Mucoromycota species in our dataset, have a *dpo* homolog; however, considering that we

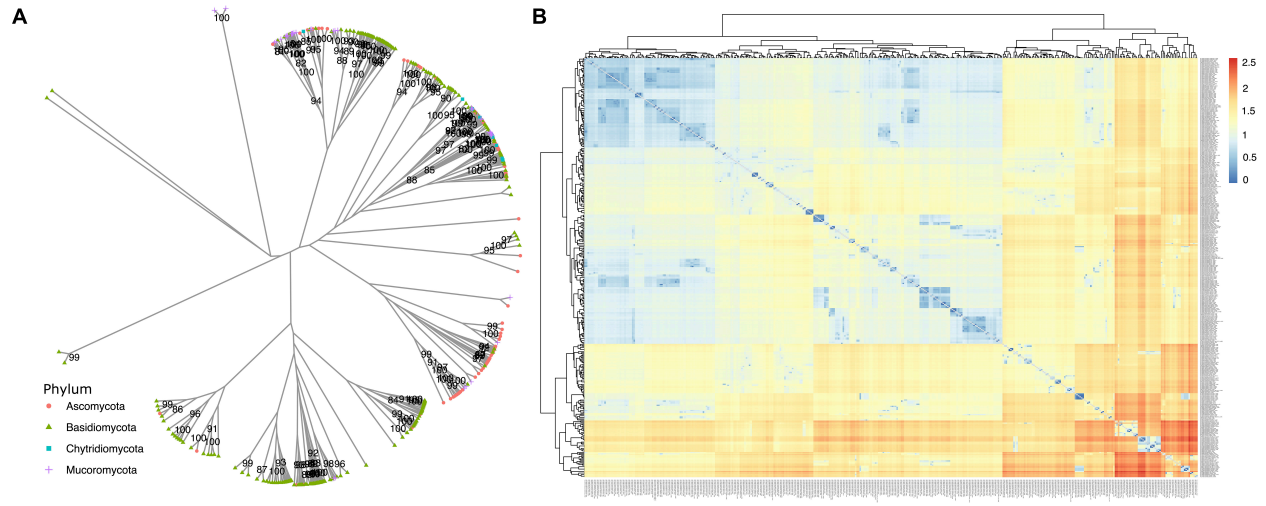


FIGURE 7 | Phylogenetic analysis of *dpo* genes. **(A)** The phylogeny was constructed by the maximum likelihood method with 1,000 bootstrap replicates. Genes are classified by phylum. **(B)** A similarity matrix based on patristic distance. Dendrograms on top and left group related mitogenomes and distances are shown from closer mitogenomes (blue) to more distant ones (red).

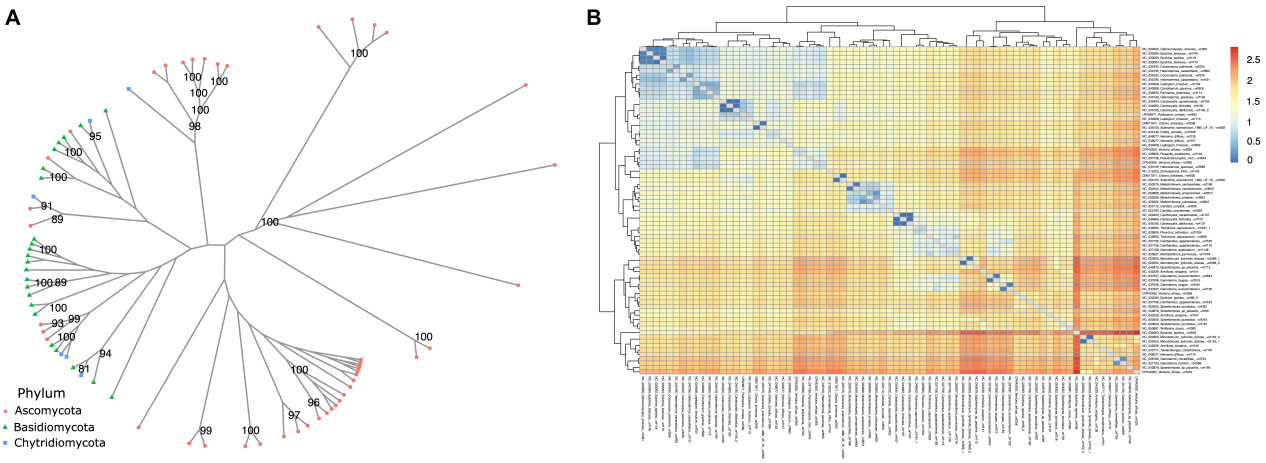


FIGURE 8 | Phylogenetic analysis of *rpo* genes. **(A)** The phylogeny was constructed by the maximum likelihood method with 1,000 bootstrap replicates. Genes are classified by phylum. **(B)** A similarity matrix based on patristic distance. Dendrograms on top and left group related mitogenomes and distances are shown from closer mitogenomes (blue) to more distant ones (red).

are working with only 24 Mucoromycota species, it is not a fair comparison with the most represented phyla Ascomycota and Basidiomycota.

When analyzing similarity among annotated *dpo* sequences (**Figure 7**), some form phyla-specific clades, whereas others seem to be shared among species of different phyla. A similar pattern was observed for *rpo* homologs (**Figure 8**). One possible explanation for this finding is horizontal gene transfer (HGT). As discussed by Beaudet et al. (2013), mitochondrial plasmids of fungi and plants exhibit many similarities, indicating that HGT events between organisms of those kingdoms may occur (Beaudet et al., 2013). According to Handa (2008) a plasmid from *Brassica* has an ORF with similarity to *rpo* from several organisms, including fungi (Handa, 2008). Moreover, since different species of fungi often share the same habitat, the hypothesis that HGT events between fungi occur is often discussed by many studies that identified highly similar *dpo* and/or *rpo* homologs among phylogenetically distant fungi (Wang et al., 2008; Xavier et al., 2012; Andrade and Goes-Neto, 2015).

Fungal Mitogenome Conformation and Genetic Code Usage

Although most mitogenomes are circular (Kulik et al., 2021), some genomes were annotated as linear. We wondered whether the conformation of the genome would affect its length, so we compared these values using Wilcoxon's statistics and found no statistical difference (**Figure 9A**). Even though, the mitogenomes conformation had no effect in mitogenome length, fungal linear

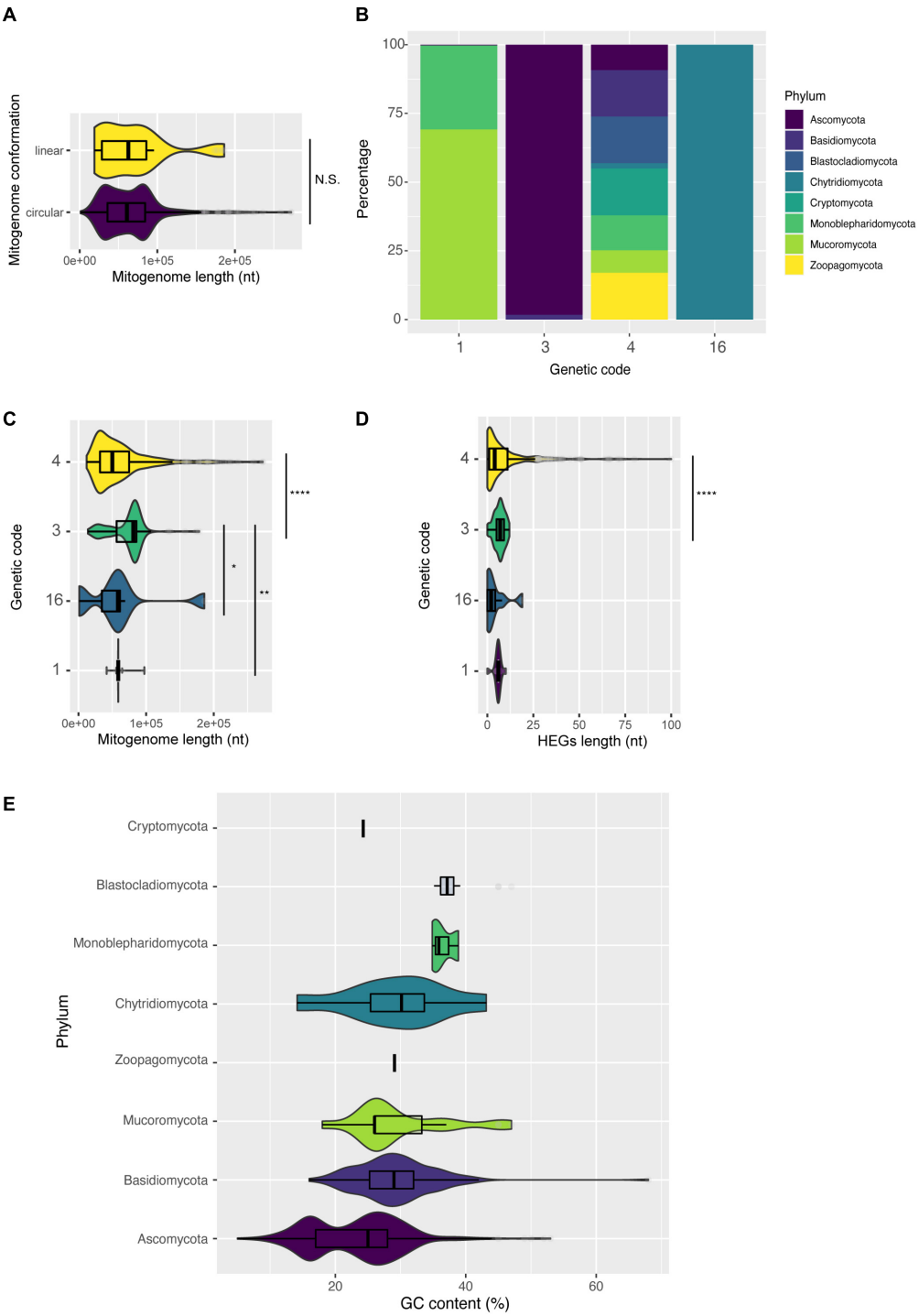


FIGURE 9 | Characterization of mitogenomes according to intrinsic molecular characteristics of the mitogenome. **(A)** Length of mitogenomes by conformation: linear or circular. **(B)** Different genetic codes identified by phylum. **(C)** Length of mitogenomes by identified genetic code. **(D)** Number of HEGs according to the genetic code, and **(E)** variation of GC content among fungal phyla.

mitogenomes have distinct features when compared with circular mitogenomes, such as the presence of an invertron (terminal inverted repeats) and of one to six uORFs that may produce DNA or RNA polymerases (Handa, 2008). In our analysis, no invertron were detected in mitogenomes. The presence of linear mitogenomes has already been described in different eukaryotes, including fungi. One of the hypotheses is that linearization occurs due to the integration of plasmids containing genes that code for

dpo (Williamson, 2002; van de Vossenberg et al., 2018). In our analysis, 37 (5.22%) linear genomes were identified and seven of them have *dpo* homologs. Examples are *Candida subhashii* and *Saccharomycopsi malanga,* with two *dpo* each. Despite the presence of these genes in some linear mitogenomes, the vast majority did not present *dpo*, suggesting that other mechanisms may be involved in mitogenome linearization.

Another little explored feature is the usage of different genetic codes in fungal mitogenomes. Among the 685 mitogenomes evaluated, the use of four different genetic codes was detected: code 1 (The Standard Code), code 3 (The Yeast Mitochondrial Code), code 4 (The Mold, Protozoan, and Coelenterate Mitochondrial Code and the *Mycoplasma*/*Spiroplasma* Code) and code 16 (Chlorophycean Mitochondrial Code)[6]. Overall, 62% of all mitogenomes use genetic code 4 (the UGA codon codes for Tryptophan), belonging to Ascomycota, Basidiomycota, Monoblepharidomycota, Cryptomycota, and Blastocladiomycota (**Figure 9B**). Most species of Mucoromycota are coded by code 1, while many species of Chytridiomycota are coded by code 16 (the TAG codon codes for Leucine) (**Figure 9B** and **Supplementary Table 7**). These results defy the classical 'Frozen Accident Theory' that proposes the immutability of the standard genetic code to avoid lethal genetic alterations (Crick, 1968). According to our genetic table analysis, most fungal species likely use the UGA codon (stop codon in standard genetic code 1) for translating Trp, similar to what was previously reported for UAA (Knight et al., 2001). Nonetheless, there is evidence that, in fact, a minority of fungal mitochondria uses UGA to code Tryptophan, showing that the annotation of fungal mitogenomes needs to be widely updated (Nibert, 2017). There are multiple theories to explain the differences of genetic code usage, but the one that better fits the fungal scenario may be related to the simplification of the mitochondrial genome and translation machinery ('genome streamlining' hypothesis), that would make the replication of mitogenomes easier, less energy costly and would also slowly simplify the mitogenome length (Andersson and Kurland, 1995; Knight et al., 2001; Swire et al., 2005).

We then compared the length of mitogenomes with different genetic codes (**Figure 9C**). Our results showed that fungi with genetic code 3 have larger mitogenomes than fungi with other genetic codes. According to NCBI Taxonomy[7], the genetic code 3 classically represents yeasts. These unicellular fungi do not necessarily need mitogenomes for energy production, and they might have more accessory elements or non-coding regions, which contributes to the genomic size increase (Ruan et al., 2017). Many studies have demonstrated that the length of mitogenomes is generally affected by the presence of HEGs (Fonseca et al., 2020; Araújo et al., 2021), and we detected the same trend in our data when comparing genomes with genetic code 3 and 4 (**Figure 9D**).

We also observed a difference between the GC content of mitogenomes at phylum level (**Figure 9E**). Usually, fungal mitogenomes have a low GC content, and in our dataset the mean GC content was ∼25%. Ascomycota, Mucoromycota and Basidiomycota showed the highest variation. Some outliers can also be observed regarding GC content, such as *M. furfur* (Basidiomycota – CP046241) with 68% GC, some species of the genus *Candida* ranging from 13–53% (Ascomycota – NC_014337 and NC_022174), and the species *Glomus cerebriforme* with 47% GC (Mucoromycota – NC_022144). Some studies suggest that low GC content restricts mutations in mitochondrial genes (Grivell, 1989; Lang et al., 1999). Thus, mitogenomes that have a high GC content may be subject to higher mutational variations.

Overrepresented Species and Global Network Analysis

Our findings indicated that the size of mitogenomes is associated with both the presence of accessory elements and protein-coding genes. Additionally, we observed that many mitogenomes from the same species (for instance, 29% of the analyzed mitogenomes are from 19 different species) remained after the similarity filter performed, and many do not have the 14 protein-coding and ribosomal core genes (74% of the mitogenomes from the same species lack at least one core gene). These aspects limited the use of a phylogenetic analysis exploring the evolutionary relationships among species, since there were mitogenomes without any gene in common to be aligned. Thus, we performed similarity network analyses to investigate the relationships between the evaluated mitogenomes.

The first analysis was done using only species that had more than three different mitogenomes. Divergent mitogenomes were found in more than one phylum. Three main groups of mitogenomes can be cited, those from the species of the genus *Saccharomyces*, *Fusarium,* and *Aspergillus*. The genus *Saccharomyces* is one group of fungi that exhibited more than one divergent mitogenome, which after the 90% similarity filter, retained 149 mitogenomes (out of 166 mitogenomes – 10.24% reduction). *Saccharomyces* can produce energy via the functioning of the mitochondrial organelle or fermentation. Therefore, these organisms can survive only with fermentable carbon sources and even without the presence of a mitogenome. These yeasts present a different phenotype, described as "petit format," unable to grow on non-fermentable carbon sources (Bulder, 1964; Goldring et al., 1970; Day, 2013) and their mitogenomes may suffer a slightly different selective pressure in relation to other groups of fungi that depend on mitochondria to grow. Furthermore, the length of mitogenomes of *Saccharomyces* can vary between two- and four-fold (Borneman and Pretorius, 2015), and this variation may be related to the expansion of intergenic regions already described in this group of organisms (Ruan et al., 2017).

Another species that also had divergent mitogenomes is *Aspergillus flavus*, for which nine mitogenomes remained after redundancy-filtering, with mitogenomes varied between 29 and 31 kbp. Similarly, the dimorphic *Cryptococcus neoformans* continued with six mitogenomes and size between 24 and 30 kbp. The species *Fusarium oxysporum* presented ten mitogenomes and sizes ranging from 34 to 52 kbp. Through the sequence similarity network, we can observe the formation of three main clusters (**Figure 10**). The first cluster, composed of mitogenomes

[6]https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi

[7]https://www.ncbi.nlm.nih.gov/taxonomy

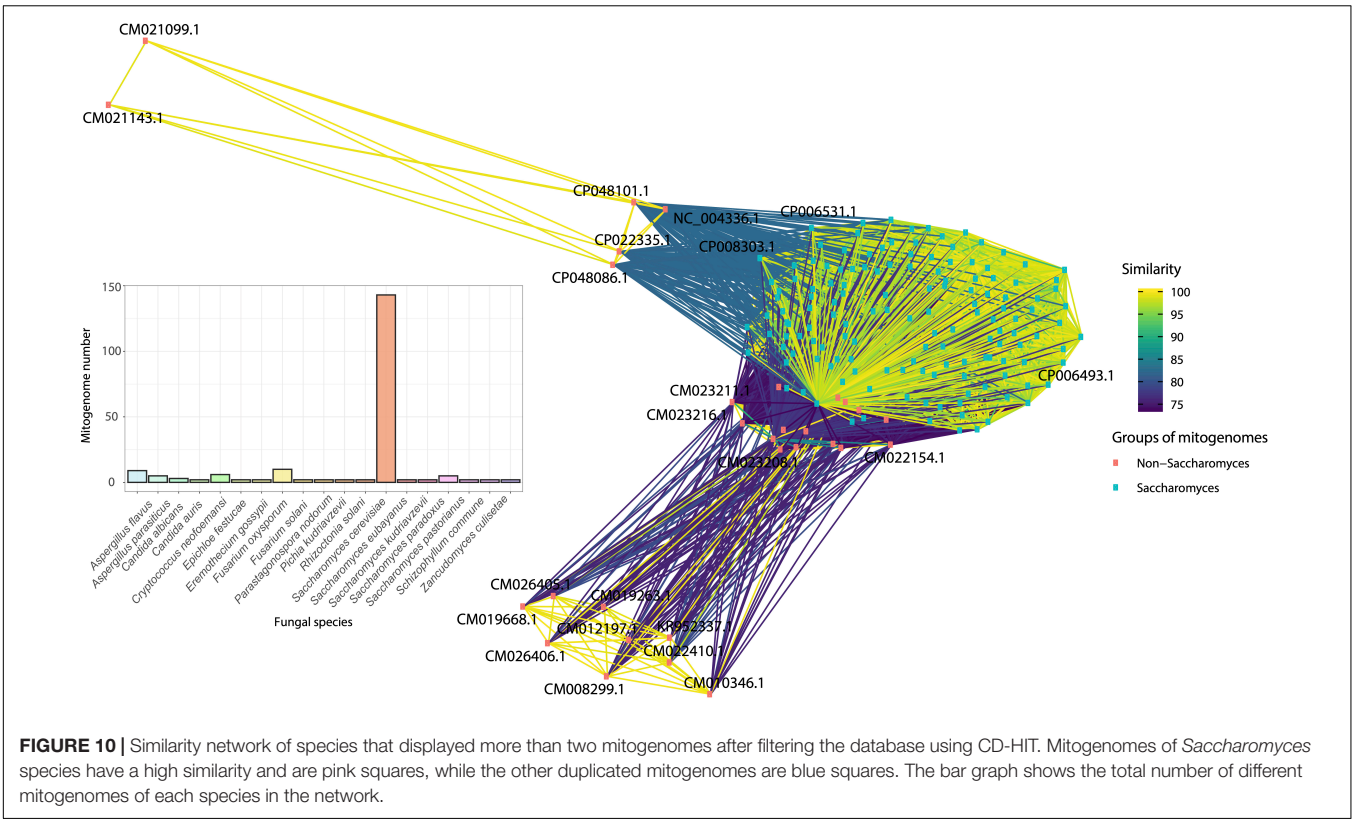


FIGURE 10 | Similarity network of species that displayed more than two mitogenomes after filtering the database using CD-HIT. Mitogenomes of *Saccharomyces* species have a high similarity and are pink squares, while the other duplicated mitogenomes are blue squares. The bar graph shows the total number of different mitogenomes of each species in the network.

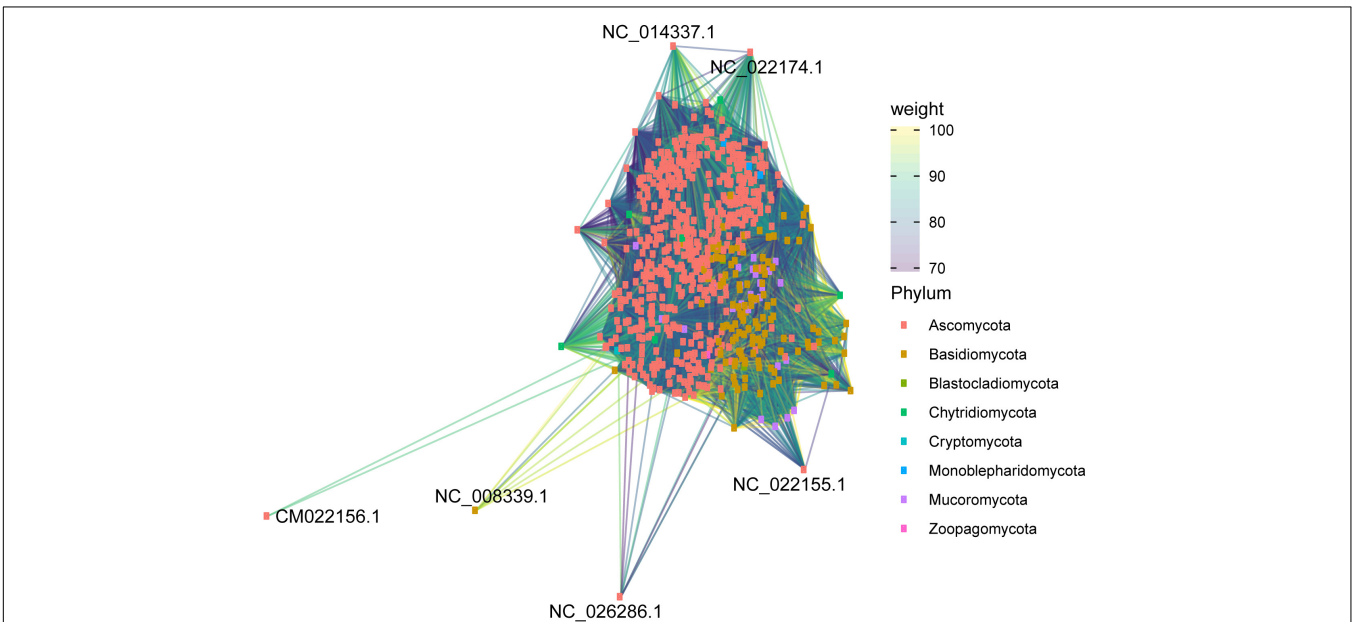


FIGURE 11 | Similarity network of all mitogenomes evaluated in this study. Mitogenomes are classified by fungal phylum.

of the genus *Saccharomyces* (green dots on bottom), the second one containing mitogenomes only from the species *F. oxysporum* (orange dots on top right) and the third, containing only the species *C. neoformans* (orange dots on top left). In this last group, it was possible to observe that some mitogenomes of *C. neoformans* (CP048101, CP048086, CP022335, and NC_004346) were grouped with a closer proximity to mitogenomes of the genus *Aspergillus* (*A. flavus* and *Aspergillus parasiticus*). In general, no significant difference in mitogenome length was observed and all *Cryptococcus* mitogenomes present

all core genes. A mitogenome of *F. oxysporum* (NC_017930) also grouped with the *Aspergillus* genus cluster. This mitogenome is the smallest of all *F. oxysporum* available (34 kbp). Size variation and differences in gene composition in the *F. oxysporum* species has been described and are associated with the presence of intronic sequences in the *nad5*, *cob* and *atp6*. Indeed, Brankovics et al. (2017), classified *F. oxysporum* strains into clades according to the presence of these sequences (Brankovics et al., 2017). The same observation was made for species of the genera *Aspergillus* and *Cryptococcus*, in which the presence of accessory elements such as introns and uORFs are responsible for the increase in size of the mitogenome (Joardar et al., 2012; Wang and Xu, 2020).

The second network analysis considered all mitogenomes evaluated to assess the grouping according to phylum-level classification. The clustering patterns among the evaluated mitogenomes are displayed in **Figure 11**. Many species were grouped according to their respective phylum, indicating that the variation in the composition of mitogenomes is lower than the similarity between the taxonomic groups of fungi. The abundance of Ascomycota and Basidiomycota mitogenomes in relation to the other phyla can also be observed. In some cases, we found clustering among different phyla, such as Ascomycota and Mucoromycota. This grouping may be due to the presence of accessory elements that can be shared between species from different phyla. In this study, phylogenetic analysis indicated that *dpo* homologs were shared between Ascomycota and Mucoromycota species (**Figure 8**), confirming that these elements may also be responsible for the grouping of mitogenomes. Other exceptions are also presented. The species *Alternaria alternata* (Ascomycota – CM022156) was distant from the other mitogenomes of the same phylum. According to our annotation, this species has 23 uORFs and only the ribosomal genes *rns* and *rnl*, which may be one of the reasons for the distance among this mitogenome and the others from Kingdom Fungi. Furthermore, Chytridiomycota representatives were not well grouped. The Chytridiomycota are pathogenic and saprotrophic fungi commonly found in freshwater environments. This phylum presents some systematic controversies about the permanence of some species in the group (Levin, 2013), and this fact may help explain the lack of grouping among the evaluated mitogenomes.

CONCLUSION AND PERSPECTIVES: WHY TO STUDY FUNGAL MITOGENOMES?

Studies that fully characterize the structure of fungal mitogenomes are still scarce. In the current study, we explored the diversity of genome structure of fungal mitogenomes at kingdom level and demonstrated that less than 0.02% of the already described fungal species have their mitogenome sequenced and publicly available. The phylum Ascomycota has the highest number of mitogenomes available, while some other phyla do not have even one single mitogenome sequenced. Taken together, our findings suggest that fungal mitogenomes exhibit high structural variation (composition, genetic codes, size, and gene diversity). The size and composition of mitogenomes were variable, mainly explained by the presence of accessory elements. Furthermore, there was no universal gene found in all fungal species. Some studies have already demonstrated the possibility of using mitochondrial genes to identify fungal species (Kulik et al., 2015; Smith, 2015; Avin et al., 2017). Nevertheless, our work indicates that there is no universal marker, which does not prevent the use of mitochondrial genes to identify groups of fungi at the order or family levels.

Our findings clearly show that, despite being considered conserved, fungal mitogenomes have numerous differences that need to be better characterized and understood. Because of this structural diversity, the use of mitogenomes are a promising tool for elucidating evolutionary relationships between species as they have a low rate of recombination, diverse patterns of mitochondrial inheritance, difference in mutation rate in relation to the nuclear genome, and facility to be amplified by PCR to sequence a particular region or the entire mitogenome due to the high number of copies in the cellular cytoplasm (Rubinoff and Holland, 2005).

Despite the important role played by mitochondria in the cell, detailed knowledge about the gene regulation of this organelle is still scarce, and very little is known about the control of expansion/contraction of mitochondrial genomes in fungi (Dirick et al., 2014; Nguyen et al., 2020; Kulik et al., 2021). The sequencing and characterization of new fungal mitochondrial genomes coupled with transcriptomic studies are of great importance to help understand genome function and gene regulation of this organelle (Grahl et al., 2012; Mosbach et al., 2017; Tang et al., 2018). Therefore, the deep characterization of fungal mitogenomes at kingdom level presented here provides a first step toward these goals.

In a more applied approach, the mitochondrial genome may also provide new strategies for controlling fungal pathogens, as well as may help improve the production of secondary metabolites (Grahl et al., 2012; Mosbach et al., 2017; Tang et al., 2018), since fungal mitochondria participate in several processes of pathogenicity, virulence, and drug resistance. Fungicides usually target proteins located in the mitochondrial inner membrane or mitogenome (Song et al., 2020). Furthermore, these studies still help investigate the function of mitochondria in cellular processes that are essential for the growth and development across fungi. To perform the characterization of resistance mutations, it is necessary to sequence and make available numerous mitogenomes from the same species.

We detected a large amount of mitogenomes available for Saccharomycetaceae, showing a representative and non-redundant dataset. This genetic variation in mitogenomes and phenotypic diversity observed especially in *Saccharomyces* makes the genus a good model for studies of comparative mitogenomics. Currently, most of the knowledge on comparative genomics in fungi comes from studies within Saccharomycetaceae (Freel et al., 2015; Peter et al., 2018), but mitogenomes are still not explored in populational, ecological, and evolutionary studies.

We also identified many unusual genes in mitogenomes, such as *dpo*, *rpo*, and mitochondrial genes usually located in the nuclear genome. This gene diversity may be related to HGT between species and even mitogenome length

(Rosewich and Kistler, 2000; Joardar et al., 2012; Kolesnikova et al., 2019; Fonseca et al., 2020; Araújo et al., 2021). It is even more notable when we evaluate different mitogenomes of the same species with different uORFs and genes, as observed for *Fusarium* and *Saccharomyces* species, indicating that fungal mitogenomes have a great plasticity. The implications of the presence of these genes are still poorly understood. Until December 2020, few fungal species had more than one sequenced mitogenome, preventing intraspecific comparisons. This scarcity of genomic data hinders the search for new antifungal agents and the characterization of resistance mutations, as well as the understanding of fungal infections in different hosts. As a summary, the results found in this study show that fungal mitogenomes are highly diverse, without the presence of any universally conserved gene, and vary widely in accessory elements, suggesting that they are in constant and fast change.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors. All scripts used in our study are available at: github.com/LBMCF/fungal_mitogenomes_frontiers.

AUTHOR CONTRIBUTIONS

PF, RD-P, DA, LT, TM-P, WR, L-ED-B, EA, and AG-N analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. AG-N receives a research grant for scientific productivity from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (no. 310764/2016-5).

ACKNOWLEDGMENTS

We would like to thank the Graduate Programs of Genetics (http://pggenetica.icb.ufmg.br/), Microbiology (http://www.microbiologia.icb.ufmg.br/pos/), and Bioinformatics (http://www.pgbioinfo.icb.ufmg.br/) of the Universidade Federal de Minas Gerais (UFMG), and the Graduate Program in Genetics and Molecular Biology of Universidade Estadual de Santa Cruz (https://nbcgib.uesc.br/genetica/). We also thank Luciano Rezende Fiuza Branco and Dener Eduardo Bortolini for helping with scripts for data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.787283/full#supplementary-material

Supplementary Table 1 | Overview of studies assessing fungal mitogenomes published from 1990 to 2020. Title and year of publication are shown for the 481 studies.

Supplementary Table 2 | Number of mitogenomes filtered with CD-HIT at thresholds of 99–90, 80 and 70% of identity. The reduction value in regarding the total number of mitogenomes is also shown.

Supplementary Table 3 | Detailed description of fungal mitogenomes used in this study. Accession numbers, taxonomy, mitogenome size, number of accessory elements and genome conformation are available.

Supplementary Table 4 | Overview of introns and HEGs identified in the mitogenomes assessed in this study. Introns that have domains of HEGs are also shown.

Supplementary Table 5 | Presence of conserved domains in uORFs sequences. The analysis was performed on the CDD database with standard parameters.

Supplementary Table 6 | Coding potential of uORFs. The coding potential was calculated with the CPC2 program with standard parameters.

Supplementary Table 7 | Differences between the use of codons and amino acids according to the four genetic codes identified in fungal mitogenomes.

Supplementary File 1 | *de novo* annotation files of the 685 mitogenomes investigated in this study. Annotation was performed with the Mfannot program.

Supplementary File 2 | Domains annotated in the uORF sequences of the *Malassezia furfur* mitogenome. The analysis was performed using the CDD database.

REFERENCES

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*. Budapest: BN Petrov; F Csaki, 267–281.

Al-Reedy, R. M., Malireddy, R., Dillman, C. B., and Kennell, J. C. (2012). Comparative analysis of Fusarium mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genet. Biol.* 49, 2–14. doi: 10.1016/j.fgb.2011.11.008

Andersson, S. G., and Kurland, C. G. (1995). Genomic evolution drives the evolution of the translation system. *Biochem. Cell Biol.* 73, 775–787. doi: 10.1139/o95-086

Andrade, B. S., and Goes-Neto, A. (2015). Phylogenetic analysis of DNA and RNA polymerases from a Moniliophthora perniciosa mitochondrial plasmid reveals probable lateral gene transfer. *Genet. Mol. Res.* 14, 14105–14114. doi: 10.4238/2015.October.29.30

Araújo, D. S., De-Paula, R. B., Tomé, L. M. R., Quintanilha-Peixoto, G., Salvador-Montoya, C. A., Del-Bem, L.-E., et al. (2021). Comparative mitogenomics of Agaricomycetes: diversity, abundance, impact and coding potential of putative open-reading frames. *Mitochondrion* 58, 1–13. doi: 10.1016/j.mito.2021.02.002

Arbuthnot, P. (2015). "Engineering sequence-specific DNA binding proteins for antiviral gene editing," in *Gene Therapy for Viral Infections*, ed. P. Arbuthnot (Amsterdam: Elsevier), 63–94.

Avin, F. A., Subha, B., Tan, Y.-S., Braukmann, T. W. A., Vikineswary, S., and Hebert, P. D. N. (2017). Escaping introns in COI through cDNA barcoding of

mushrooms: pleurotus as a test case. *Ecol. Evol.* 7, 6972–6980. doi: 10.1002/ece3.3049

Barzel, A., Obolski, U., Gogarten, J., Kupiec, M., and Hadany, L. (2011). Home and away- the evolutionary dynamics of homing endonucleases. *BMC Evol. Biol.* 11:324. doi: 10.1186/1471-2148-11-324

Beaudet, D., Terrat, Y., Halary, S., de la Providencia, I. E., and Hijri, M. (2013). Mitochondrial genome rearrangements in glomus species triggered by homologous recombination between distinct mtDNA haplotypes. *Genome Biol. Evol.* 5, 1628–1643. doi: 10.1093/gbe/evt120

Belfort, M. (2003). Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev.* 17, 2860–2863. doi: 10.1101/gad.1162503

Berg, O. G., and Kurland, C. G. (2000). Why mitochondrial genes are most often found in nuclei. *Mol. Biol. Evol.* 17, 951–961. doi: 10.1093/oxfordjournals.molbev.a026376

Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., and Pfanner, N. (2008). Multiple pathways for sorting mitochondrial precursor proteins. *EMBO Rep.* 9, 42–49. doi: 10.1038/sj.embor.7401126

Borneman, A. R., and Pretorius, I. S. (2015). Genomic insights into the saccharomyces sensu stricto complex. *Genetics* 199, 281–291. doi: 10.1534/genetics.114.173633

Brankovics, B., van Dam, P., Rep, M., de Hoog, G. S., van der Lee, T. A., Waalwijk, C., et al. (2017). Mitochondrial genomes reveal recombination in the presumed asexual Fusarium oxysporum species complex. *BMC Genomics* 18:735. doi: 10.1186/s12864-017-4116-5

Bulder, C. J. E. A. (1964). Induction of petite mutation and inhibition of synthesis of respiratory enzymes in various yeasts. *Antonie van Leeuwenhoek* 30, 1–9. doi: 10.1007/BF02046695

Bullerwell, C. E. (2003). A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res.* 31, 759–768. doi: 10.1093/nar/gkg134

Carvalho, D. S., Andrade, R. F. S., Pinho, S. T. R., Góes-Neto, A., Lobão, T. C. P., Bomfim, G. C., et al. (2015). What are the evolutionary origins of mitochondria? A complex network approach. *PLoS One* 10:e0134988. doi: 10.1371/journal.pone.0134988

Cheek, M., Lughadha, E. N., Kirk, P., Lindon, H., Carretero, J., Looney, B., et al. (2020). New scientific discoveries: plants and fungi. *Plants People Planet* 2, 371–388. doi: 10.1002/ppp3.10148

Chen, C., Li, Q., Fu, R., Wang, J., Xiong, C., Fan, Z., et al. (2019). Characterization of the mitochondrial genome of the pathogenic fungus Scytalidium auriculariicola (Leotiomycetes) and insights into its phylogenetics. *Sci. Rep.* 9:17447. doi: 10.1038/s41598-019-53941-5

Chevalier, B. S., and Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* 29, 3757–3774.

Crick, F. H. (1968). The origin of the genetic code. *J. Mol. Biol.* 38, 367–379. doi: 10.1016/0022-2836(68)90392-6

Day, M. (2013). Yeast petites and small colony variants. *Adv. Appl. Microbiol.* 85, 1–41. doi: 10.1016/B978-0-12-407672-3.00001-0

Dirick, L., Bendris, W., Loubiere, V., Gostan, T., Gueydon, E., and Schwob, E. (2014). Metabolic and environmental conditions determine nuclear genomic instability in budding yeast lacking mitochondrial DNA. *G3 Genes Genomes Genet.* 4, 411–423. doi: 10.1534/g3.113.010108

Dujon, B. (2010). Yeast evolutionary genomics. *Nat. Rev. Genet.* 11, 512–524. doi: 10.1038/nrg2811

Edgell, D. R. (2009). Selfish DNA: homing endonucleases find a home. *Curr. Biol.* 19, R115–R117. doi: 10.1016/j.cub.2008.12.019

Férandon, C., Xu, J., and Barroso, G. (2013). The 135 kbp mitochondrial genome of Agaricus bisporus is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. *Fungal Genet. Biol.* 55, 85–91. doi: 10.1016/j.fgb.2013.01.009

Fonseca, P. L. C., Badotti, F., De-Paula, R. B., Araújo, D. S., Bortolini, D. E., Del-Bem, L.-E., et al. (2020). Exploring the relationship among divergence time and coding and non-coding elements in the shaping of fungal mitochondrial genomes. *Front. Microbiol.* 11:765. doi: 10.3389/fmicb.2020.00765

Formaggioni, A., Luchetti, A., and Plazzi, F. (2021). Mitochondrial genomic landscape: a portrait of the mitochondrial genome 40 years after the first complete sequence. *Life* 11:663. doi: 10.3390/life11070663

Formighieri, E. F., Tiburcio, R. A., Armas, E. D., Medrano, F. J., Shimo, H., Carels, N., et al. (2008). The mitochondrial genome of the phytopathogenic basidiomycete Moniliophthora perniciosa is 109kb in size and contains a stable integrated plasmid. *Mycol. Res.* 112, 1136–1152. doi: 10.1016/j.mycres.2008.04.014

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

Goldring, E. S., Grossman, L. I., Krupnick, D., Cryer, D. R., and Marmur, J. (1970). The petite mutation in Yeast. *J. Mol. Biol.* 52, 323–335. doi: 10.1016/0022-2836(70)90033-1

Grahl, N., Dinamarco, T. M., Willger, S. D., Goldman, G. H., and Cramer, R. A. (2012). Aspergillus fumigatus mitochondrial electron transport chain mediates oxidative stress homeostasis, hypoxia responses and fungal pathogenesis: mitochondrial respiration and Aspergillus pathogenesis. *Mol. Microbiol.* 84, 383–399. doi: 10.1111/j.1365-2958.2012.08034.x

Grivell, L. A. (1989). Nucleo-mitochondrial interactions in yeast mitochondrial biogenesis. *Eur. J. Biochem.* 182, 477–493. doi: 10.1111/j.1432-1033.1989.tb14854.x

Hafez, M., and Hausner, G. (2012). Homing endonucleases: DNA scissors on a mission. *Genome* 55, 553–569. doi: 10.1139/g2012-049

Handa, H. (2008). Linear plasmids in plant mitochondria: peaceful coexistences or malicious invasions? *Mitochondrion* 8, 15–25. doi: 10.1016/j.mito.2007.10.002

Hausner, G. (2003). "Fungal mitochondrial genomes, plasmids and introns," in *Applied Mycology and Biotechnology*, eds D. K. Arora and G. G. Khachatourians (Amsterdam: Elsevier), 101–131.

Hausner, G., Hafez, M., and Edgell, D. R. (2014). Bacterial group I introns: mobile RNA catalysts. *Mobile DNA* 5:8. doi: 10.1186/1759-8753-5-8

James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., et al. (2013). Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr. Biol.* 23, 1548–1553. doi: 10.1016/j.cub.2013.06.057

James, T. Y., Stajich, J. E., Hittinger, C. T., and Rokas, A. (2020). Toward a fully resolved fungal tree of life. *Annu. Rev. Microbiol.* 74, 291–313. doi: 10.1146/annurev-micro-022020-051835

Joardar, V., Abrams, N. F., Hostetler, J., Paukstelis, P. J., Pakala, S., Pakala, S. B., et al. (2012). Sequencing of mitochondrial genomes of nine Aspergillus and Penicillium species identifies mobile introns and accessory genes as main sources of genome size variability. *BMC Genomics* 13:698. doi: 10.1186/1471-2164-13-698

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201

Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kaniak-Golik, A., and Skoneczna, A. (2015). Mitochondria–nucleus network for genome stability. *Free Radical Biol. Med.* 82, 73–104. doi: 10.1016/j.freeradbiomed.2015.01.013

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Kirk, P. (2020). *Species Fungorum for CoL+. Catalogue of Life Checklist*. Kew: Royal Botanic Gardens.

Knight, R. D., Freeland, S. J., and Landweber, L. F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol.* 2:RESEARCH0010. doi: 10.1186/gb-2001-2-4-research0010

Kolesnikova, A. I., Putintseva, Y. A., Simonov, E. P., Biriukov, V. V., Oreshkova, N. V., Pavlov, I. N., et al. (2019). Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related Armillaria species. *BMC Genomics* 20:351. doi: 10.1186/s12864-019-5732-z

Kulik, T., Ostrowska, A., Buśko, M., Pasquali, M., Beyer, M., Stenglein, S., et al. (2015). Development of an FgMito assay: a highly sensitive mitochondrial based

qPCR assay for quantification of Fusarium graminearum sensu stricto. *Int. J. Food Microbiol.* 210, 16–23. doi: 10.1016/j.ijfoodmicro.2015.06.012

Kulik, T., Van Diepeningen, A. D., and Hausner, G. (2021). Editorial: the significance of mitogenomics in mycology. *Front. Microbiol.* 11:628579. doi: 10.3389/fmicb.2020.628579

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lambowitz, A. M., and Perlman, P. S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem. Sci.* 15, 440–444. doi: 10.1016/0968-0004(90)90283-H

Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3:a003616. doi: 10.1101/cshperspect.a003616

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397. doi: 10.1146/annurev.genet.33.1.351

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

Lazarus, K. L., and James, T. Y. (2015). Surveying the biodiversity of the Cryptomycota using a targeted PCR approach. *Fungal Ecol.* 14, 62–70. doi: 10.1016/j.funeco.2014.11.004

Letcher, P. M., and Powell, M. J. (2018). A taxonomic summary and revision of Rozella (Cryptomycota). *IMA Fungus* 9, 383–399. doi: 10.5598/imafungus.2018.09.02.09

Levin, S. A. (2013). *Encyclopedia of Biodiversity*, 2nd Edn. Amsterdam: Elsevier.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Li, Y., Steenwyk, J. L., Chang, Y., Wang, Y., James, T. Y., Stajich, J. E., et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Curr. Biol.* 31, 1653.e5–1665.e5. doi: 10.1016/j.cub.2021.01.074

Liu, W., Cai, Y., Zhang, Q., Shu, F., Chen, L., Ma, X., et al. (2020). Subchromosome-Scale nuclear and complete mitochondrial genome characteristics of morchella crassipes. *Int. J. Mol. Sci.* 21:483. doi: 10.3390/ijms21020483

Loizos, N., Tillier, E. R., and Belfort, M. (1994). Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11983–11987. doi: 10.1073/pnas.91.25.11983

Megarioti, A. H., and Kouvelis, V. N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.* 12, 1337–1354. doi: 10.1093/gbe/evaa126

Mendoza, H., Perlin, M. H., and Schirawski, J. (2020). Mitochondrial inheritance in phytopathogenic fungi—everything is known, or is it? *Int. J. Mol. Sci.* 21:3883. doi: 10.3390/ijms21113883

Mosbach, A., Edel, D., Farmer, A. D., Widdison, S., Barchietto, T., Dietrich, R. A., et al. (2017). Anilinopyrimidine resistance in botrytis cinerea is linked to mitochondrial function. *Front. Microbiol.* 8:2361. doi: 10.3389/fmicb.2017.02361

Muñoz-Gómez, S. A., Wideman, J. G., Roger, A. J., and Slamovits, C. H. (2017). The origin of mitochondrial cristae from alphaproteobacteria. *Mol. Biol. Evol.* 34, 943–956. doi: 10.1093/molbev/msw298

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. doi: 10.1186/s12862-020-01685-6

Nibert, M. L. (2017). Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. *Virology* 507, 96–100. doi: 10.1016/j.virol.2017.04.010

Paquin, B., Laforest, M.-J., Forget, L., Roewer, I., Wang, Z., Longcore, J., et al. (1997). The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr. Genet.* 31, 380–395. doi: 10.1007/s002940050220

Pedersen, T. L., and RStudio (2021). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. Available online at: https://CRAN.R-project.org/package=ggraph (accessed September 17, 2021).

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Robbertse, B., Reeves, J. B., Schoch, C. L., and Spatafora, J. W. (2006). A phylogenomic analysis of the Ascomycota. *Fungal Genetics Biol.* 43, 715–725. doi: 10.1016/j.fgb.2006.05.001

Rosewich, U. L., and Kistler, H. C. (2000). Role of horizontal gene transfer in the evolution of fungi. *Annu. Rev. Phytopathol.* 38, 325–363. doi: 10.1146/annurev.phyto.38.1.325

Ruan, J., Cheng, J., Zhang, T., and Jiang, H. (2017). Mitochondrial genome evolution in the Saccharomyces sensu stricto complex. *PLoS One* 12:e0183035. doi: 10.1371/journal.pone.0183035

Rubinoff, D., and Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst. Biol.* 54, 952–961. doi: 10.1080/10635150500234674

Sankaranarayanan, S. R., Ianiri, G., Coelho, M. A., Reza, M. H., Thimmappa, B. C., Ganguly, P., et al. (2020). Loss of centromere function drives karyotype evolution in closely related Malassezia species. *eLife* 9:e53944. doi: 10.7554/eLife.53944

Schmidt-Dannert, C. (2016). Biocatalytic portfolio of Basidiomycota. *Curr. Opin. Chem. Biol.* 31, 40–49. doi: 10.1016/j.cbpa.2016.01.002

Schoch, C. L., Sung, G.-H., López-Giráldez, F., Townsend, J. P., Miadlikowska, J., Hofstetter, V., et al. (2009). The ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Syst. Biol.* 58, 224–239. doi: 10.1093/sysbio/syp020

Shafer, K. S., Hanekamp, T., White, K. H., and Thorsness, P. E. (1999). Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae. *Curr. Genet.* 36, 183–194. doi: 10.1007/s002940050489

Smith, D. R. (2015). The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Briefings Funct. Genomics* 15, 47–54. doi: 10.1093/bfgp/elv027

Song, N., Geng, Y., and Li, X. (2020). The mitochondrial genome of the phytopathogenic fungus bipolaris sorokiniana and the utility of mitochondrial genome to infer phylogeny of dothideomycetes. *Front. Microbiol.* 11:863. doi: 10.3389/fmicb.2020.00863

Spinelli, J. B., and Haigis, M. C. (2018). The multifaceted contributions of mitochondria to cellular metabolism. *Nat. Cell Biol.* 20, 745–754. doi: 10.1038/s41556-018-0124-1

Stoddard, B. L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* 5:7. doi: 10.1186/1759-8753-5-7

Swire, J., Judson, O. P., and Burt, A. (2005). Mitochondrial genetic codes evolve to match amino acid requirements of proteins. *J. Mol. Evol.* 60, 128–139. doi: 10.1007/s00239-004-0077-9

Tang, G., Zhang, C., Ju, Z., Zheng, S., Wen, Z., Xu, S., et al. (2018). The mitochondrial membrane protein FgLetm1 regulates mitochondrial integrity, production of endogenous reactive oxygen species and mycotoxin biosynthesis in Fusarium graminearum: function of FgLetm1 in Fusarium graminearum. *Mol. Plant Pathol.* 19, 1595–1611. doi: 10.1111/mpp.12633

Theelen, B., Cafarchia, C., Gaitanis, G., Bassukas, I. D., Boekhout, T., and Dawson, T. L. (2018). Malassezia ecology, pathophysiology, and treatment. *Med. Mycol.* 56, S10–S25. doi: 10.1093/mmy/myx134

Torriani, S. F. F., Penselin, D., Knogge, W., Felder, M., Taudien, S., Platzer, M., et al. (2014). Comparative analysis of mitochondrial genomes from closely related Rhynchosporium species reveals extensive intron invasion. *Fungal Genet. Biol.* 62, 34–42. doi: 10.1016/j.fgb.2013.11.001

Tsuji, J., Frith, M. C., Tomii, K., and Horton, P. (2012). Mammalian NUMT insertion is non-random. *Nucleic Acids Res.* 40, 9073–9088. doi: 10.1093/nar/gks424

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

Verma, S., Shakya, V. P. S., and Idnurm, A. (2018). Exploring and exploiting the connection between mitochondria and the virulence of human pathogenic fungi. *Virulence* 9, 426–446. doi: 10.1080/21505594.2017.1414133

Walker, G. M., and White, N. A. (2005). "Introduction to fungal physiology," in *Fungi*, ed. K. Kavanagh (Chichester: John Wiley & Sons, Ltd), 1–34.

Wallen, R. M., and Perlin, M. H. (2018). An overview of the function and maintenance of sexual reproduction in dikaryotic fungi. *Front. Microbiol.* 9:503. doi: 10.3389/fmicb.2018.00503

Wang, Y., and Xu, J. (2020). Mitochondrial genome polymorphisms in the human pathogenic fungus cryptococcus neoformans. *Front. Microbiol.* 11:706. doi: 10.3389/fmicb.2020.00706

Wang, Y., Zeng, F., Hon, C. C., Zhang, Y., and Leung, F. C. C. (2008). The mitochondrial genome of the Basidiomycete fungus Pleurotus ostreatus (oyster mushroom). *FEMS Microbiol. Lett.* 280, 34–41. doi: 10.1111/j.1574-6968.2007.01048.x

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer International Publishing.

Williamson, D. (2002). The curious history of yeast mitochondrial DNA. *Nat. Rev. Genet.* 3, 475–481. doi: 10.1038/nrg814

Xavier, B. B., Miao, V. P. W., Jónsson, Z. O., and Andrésson, O. S. (2012). Mitochondrial genomes from the lichenized fungi Peltigera membranacea and Peltigera malacea: features and phylogeny. *Fungal Biol.* 116, 802–814. doi: 10.1016/j.funbio.2012.04.013

Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Curr. Protocols Bioinform.* 69:e96. doi: 10.1002/cpbi.96

Zardoya, R. (2020). Recent advances in understanding mitochondrial genome diversity. *F1000Res* 9:270. doi: 10.12688/f1000research.21490.1

Conflict of Interest: The 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.

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.

Citation: Fonseca PLC, De-Paula RB, Araújo DS, Tomé LMR, Mendes-Pereira T, Rodrigues WFC, Del-Bem L-E, Aguiar ERGR and Góes-Neto A (2021) Global Characterization of Fungal Mitogenomes: New Insights on Genomic Diversity and Dynamism of Coding Genes and Accessory Elements. Front. Microbiol. 12:787283. doi: 10.3389/fmicb.2021.787283

frontiers in Fungal Biology

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/ffunb.2021.779337

Mitonuclear Genetic Interactions in the Basidiomycete *Heterobasidion parviporum* Involve a Non-conserved Mitochondrial Open Reading Frame

Pierre-Henri Clergeot† *and Åke Olson**

Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden

The mitochondrial and nuclear genomes of Eukaryotes are inherited separately and consequently follow distinct evolutionary paths. Nevertheless, the encoding of many mitochondrial proteins by the nuclear genome shows the high level of integration they have reached, which makes mitonuclear genetic interactions all the more conceivable. For each species, natural selection has fostered the evolution of coadapted alleles in both genomes, but a population-wise divergence of such alleles could lead to important phenotypic variation, and, ultimately, to speciation. In this study in the Basidiomycete *Heterobasidion parviporum*, we have investigated the genetic basis of phenotypic variation among laboratory-designed heterokaryons carrying the same pair of haploid nuclei, but a different mitochondrial genome. Radial growth rate data of thirteen unrelated homokaryotic parents and of their heterokaryotic offspring were combined with SNP data extracted from parental genome sequences to identify nuclear and mitochondrial loci involved in mitonuclear interactions. Two nuclear loci encoding mitochondrial proteins appeared as best candidates to engage in a genetic interaction affecting radial growth rate with a non-conserved mitochondrial open reading frame of unknown function and not reported apart from the Russulales order of Basidiomycete fungi. We believe our approach could be useful to investigate several important traits of fungal biology where mitonuclear interactions play a role, including virulence of fungal pathogens.

Keywords: heterokaryon, mitochondria, nucleus, mycelium, growth, cytonuclear

OPEN ACCESS

Edited by:
Vassili N. Kouvelis, National and Kapodistrian University of Athens, Greece

Reviewed by:
Timothy Yong James, University of Michigan, United States
Fabiano Sillo, Istituto per la Protezione Sostenibile Delle Piante, Italy

***Correspondence:**
Åke Olson
ake.olson@slu.se

†***Present address:***
Pierre-Henri Clergeot, FREDON Grand-Est, site de Sélestat, France

Specialty section:
This article was submitted to Fungal Genomics and Evolution, a section of the journal Frontiers in Fungal Biology

Received: *18 September 2021*
Accepted: *19 November 2021*
Published: *14 December 2021*

Citation:
Clergeot P-H and Olson Å (2021) Mitonuclear Genetic Interactions in the Basidiomycete Heterobasidion parviporum Involve a Non-conserved Mitochondrial Open Reading Frame. Front. Fungal Biol. 2:779337. doi: 10.3389/ffunb.2021.779337

INTRODUCTION

Mitochondria are functionally and genetically dynamic organelles essential to the eukaryotic cell (Burger et al., 2003; Henze and Martin, 2003; McBride et al., 2006). They are the center of aerobic production of energy, but also play a role in signaling between cellular components, cellular growth and differentiation, apoptosis, control of the cell cycle, disease, and aging (McBride et al., 2006; Bernhardt et al., 2014; Urbani and Babu, 2019). Although the majority of mitochondrial proteins are encoded by nuclear genes, mitochondria have their own self-replicating genome bearing phylogenetic evidence of prokaryotic ancestry (Gray et al., 2001). According to the endosymbiotic theory, mitochondria originate from the endocytosis of an alphaproteobacterium that took place about two billions years ago as an adaptation to an atmosphere richer in oxygen (Gray et al., 1999). Mitochondrial and nuclear genomes have subsequently coevolved throughout the evolutionary

history of Eukaryota, following two major trends: one of reduction of the number of mitochondrial genes non-essential to the eukaryotic cell, the other of transfer to the nucleus of unique essential genes whose products can be imported from the cytosol to the mitochondrion without fitness cost (Kurland and Andersson, 2000). Mitochondrial genes encoding proteins involved in cellular respiration are highly conserved while other loci are more variable depending on the species, and some contribute to mitochondrial genome evolution and plasticity, like highly polymorphic, repetitive elements found within introns and intergenic regions (Burger et al., 2003). Consequently, conformation of the mitochondrial genome, its size, its degree of relatedness to its bacterial ancestor, its genes number, their order and their expression vary by species (Gray et al., 1999; Medina et al., 2020 for a recent review in fungi). Maintaining mitochondrial activity with a reduced genome implies multiple mitonuclear genetic interactions (Stojković and Dordević, 2017), but also requires nuclear and mitochondrial DNA permanently to coadapt despite their frequently different mutation rates (Allio et al., 2017; Sandor et al., 2018; Medina et al., 2020). Population-wise divergence of mitochondrial and nuclear coadapted alleles is predicted to lead to speciation due to Bateson-Dobzhansky-Muller incompatibility (Nguyen et al., 2020 and references therein).

The goal of this article is to show that harnessing natural intraspecific variation of coadapted mitochondrial and nuclear alleles in order to study genetics of mitonuclear interactions is possible. Natural variation in the mitochondrial genome of two species of the tree pathogen *Heterobasidion* is associated with the emergence of interspecific heterokaryotic hybrids with virulence that correlates with mitochondrial inheritance (Olson and Stenlid, 2001; Giordano et al., 2018). This observation suggests that, in self-incompatible Basidiomycetes species such as *Heterobasidion spp*., new combinations of mitochondrial and nuclear DNA formed at the heterokaryotic phase could drastically alter phenotype. Isolates of *Heterobasidion spp.* spend most of their life cycle as heterokaryotic mycelium (n + n) in host wood (Johannesson and Stenlid, 2004; Garbelotto and Gonthier, 2013). The homokaryotic (haploid) phase is short and exposed mainly to purifying and neutral selections: Basidiospores dispersal by the wind, their landing on an exposed area of a potential host, and the encounter of homokaryotic mycelia originating from different basidiocarps on host wood are largely stochastic. Mating is under the control of a unifactorial multiallelic genetic system (Holt et al., 1983), which, preceded by basidiospore dispersal, is supposed to favor outbreeding (Nieuwenhuis et al., 2013). It is therefore when mitochondrial and nuclear DNA of different homokaryotic isolates meet at the heterokaryotic phase that new mitonuclear combinations might emerge. If so, some degree of mitochondrial DNA-dependent phenotypic variation is expected at each generation of intraspecific heterokaryons as well.

Our recent work in *Heterobasdion parviporum* demonstrated and quantified mitochondrial genetic effects on two fitness proxies in heterokaryons and their parents (Clergeot et al., 2019). Our experimental design was based on heterokaryon synthesis (Olson and Stenlid, 2001), during which mitochondrial DNA is unparentally inherited while nuclei undergo bidirectional exchange, resulting in a pair of heterokaryons with the same haploid nuclei but a different mitochondrial genome (Xu and Wang, 2015). Heterokaryon synthesis was implemented with thirty homokaryotic parental isolates sampled at various locations in Eurasia and showing no population structure. We observed that variance of growth rate does not correlate with genetic distance in heterokaryons and that mitochondrial genetic effects account for 35% of phenotypic variance among homokaryons and 3% among heterokaryons (Clergeot et al., 2019). Heterokaryons are more fit than their parents in average, but not more than the fittest parents, suggesting heterozygosity and the masking of deleterious alleles at some of the nuclear loci involved in mitonuclear interactions in the offspring, rather than an intrinsic difference of fitness between the heterokaryons and their parents.

The present study is an investigation of the genetic basis of mitonuclear interactions in *Heterobasidion parviporum*: As phenotyping the heterokaryotic offspring of pairwise crossings of homokaryotic parents in controlled conditions brings a more reliable assessment of the variation resulting from combining different nuclear and mitochondrial alleles (Nguyen et al., 2020), we have revisited our data of heterokaryon variance for growth rate and parental genome sequence polymorphism (Clergeot et al., 2019) in order to identify candidate loci acting in *trans* across the two genomes. We have been able to single out this way two possible candidates for a nuclear locus encoding proteins involved in molecular trafficking to the mitochondrion (a mitochondrial carrier protein or a mitochondrial calcium uniporter), and a non-conserved mitochondrial locus of unknown function and whose distribution is limited to the Russulales order of Agaricomycetes. We believe that our results on mitonuclear genetics of growth rate in *H. parviporum* could pave the way to similar investigations of mitochondrial DNA-dependent variation of important traits e.g., virulence in fungal pathogens.

MATERIALS AND METHODS

Fungal Isolates

Data from thirteen *H. parviporum* homokaryotic isolates out of thirty initially crossed in our previous study (Clergeot et al., 2019), and from 73 pairs of their heterokaryotic offspring were analyzed for mitonuclear epistasis in this study. Heterokaryons of each pair are carrying the same parental haploid nuclei, but mitochondria from either of the two parental isolates (Xu and Wang, 2015; see also **Figure 2** in Clergeot et al., 2019). All parental homokaryotic isolates were given a numeric code (crossing chart and correspondence of numeric codes to original names are available in **Supplementary Material 1**) and each isolate was labeled as follows: Homokaryotic parents are designed by their numeric code with prefix "Ho." Heterokaryons are designed by numeric codes of their parental nuclei, starting by acceptor isolate bringing the parental cytoplasm (maternal isolate), followed by donor isolate in parenthesis (paternal isolate). For example, heterokaryon 9(32) has haploid nucleus and mitochondria of homokaryotic parent Ho9, but only haploid

nucleus of homokaryotic parent Ho32. Pairs of heterokaryons originating from the same homokaryotic parents are labeled with parental numeric codes in parenthesis separated by a semicolon. For example, pair (9;32) consists of heterokaryons 9(32) and 32(9), which originate from homokaryons Ho9 and Ho32.

Phenotypic Analysis

All data of fungal mycelium growth used in this study have been published earlier in Clergeot et al. (2019). Briefly, isolates were grown on artificial medium cast in Petri dishes, and their growth was recorded daily on two or three plates, 4–6 days post inoculation, four times on each plate. Experiments were repeated at that time, although never simultaneously with all heterokaryons together for obvious practical reasons. Consequently an "assay" does not refer to a set of experiments carried out with all isolates together at the same time, but rather to a specific experiment in a series made for each individual isolate at different times between 2015 and 2017, usually two or three. Each assay for any of the 146 heterokaryons considered in this study was given a different identification number. One heterokaryon was assayed just once, 95 twice, 41 thrice, eight four times, and one six times, in a total of 352 assays, consisting in a total of 910 different plates with four values of growth rate each.

All values of radial growth rate (RGR, in mm/day) were estimated by the method of the least squares, using published data of growth measurements of parental homokaryons and their heterokaryotic offspring (see **Supplementary Material 2** gathering all growth rate data). Data dispersion was evaluated at three levels of subdivision, by comparing quartile coefficients of dispersion [QCD = (Q3-Q1)/Q3+Q1 with Q1 being the first quartile and Q3 the third quartile]. QCDs were calculated with growth rates from either each plate (910 sets of values), each assay (352 sets), or each heterokaryon (146 sets). Statistical homogeneity between data sets (i.e., heterokaryons of a same pair, or different assays of the same heterokaryon) was assessed using Mann-Whitney tests for two independent samples (two tails; $\alpha = 0.05$) with correction for ties, included in the Real Statistics Resource Pack software (Release 3.5.3). Copyright (2013–2021) Charles Zaiontz. www.real-statistics.com. Risk α for Type I error was adjusted for multiple testing using the method of Holm-Bonferroni (Holm, 1979).

Genotypic Analysis

Filtering of Mitochondrial SNPs

H. parviporum mitochondrial DNA sequence is borne by two specific unitigs, unitig_36 and unitig_37 (Clergeot et al., 2019). All SNPs present on these two unitigs were filtered as described (Clergeot et al., 2019), but using in-house Perl scripts revisited in order to include indels and to lift constraint on the quality score. All Perl scripts used in this study are available on GitHub under repository CytN (https://github.com/phcler/CytN). Positions of mitochondrial SNP loci were searched in *H. irregulare* annotated mitochondrial genome (accession NC_024555.1) using BLASTN software (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and published mitochondrial ORF predictions (Himmelstrand et al., 2014).

Filtering of Nuclear SNPs

Nuclear SNPs were filtered based on genotypic information that one allele is borne by homokaryotic isolates Ho9 and Ho32, the other borne by Ho2, Ho5, Ho11, Ho18, Ho19, Ho20, Ho27, Ho26, Ho30 and Ho31, and that allele is undetermined for all other isolates. As Ho18 was the reference isolate for SNP calling (Clergeot et al., 2019), Ho9 and Ho32 are therefore bearing the alternative allele. Apart from setting this allelic filter, from including biallelic indels and from relaxing constraint on the quality score (lowered from 10,000 to 100), all other filtering steps were performed using in-house Perl scripts as previously described (Clergeot et al., 2019). An automated procedure was designed to find out whether or not filtered SNPs are: (1) localized in the open reading frame of a nuclear gene in *H. parviporum*, (2) at the origin of a non-synonymous mutation in isolates Ho9 and Ho32. In brief, the procedure relies on standalone nucleotide blast searches with *H. parviporum* genomic fragments surrounding each SNP as queries against a custom database including 13,405 open reading frames (ORF) of the related species *H. irregulare* (Olson et al., 2012). Blast results files were subsequently analyzed with a set of in-house Perl scripts to detect non-synonymous substitutions in the coding sequence of a gene. SNPs predicted to localize in an intron, or up to 800 bp upstream/downstream of the first/last exon of a *H. parviporum* gene were analyzed with separate in-house Perl scripts. Details about these procedures of SNP filtering and analysis are provided in **Supplementary Materials 3, 4**.

Expression Analysis of Mitochondrial Locus *nc-ORF5* in *H. irregulare*

Transcripts of the 936 bp-long hypothetical ORF (positions 46939-47874 in *H. irregulare* mitochondrial genome; accession NC_024555.1) published as *nc-ORF5* in Himmelstrand et al. (2014) were searched among published RNA sequencing data of *H. irregulare* strain TC32-1 (Olson et al., 2012) using magic-BLAST software (https://ncbi.github.io/magicblast/). Proteins similar to actual *nc-ORF5* translation product were searched using BLASTP software (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

RESULTS

From the thirty homokaryotic isolates of *H. parviporum* in our previous study from 2019, reciprocal exchange of nuclei during mating was successful with thirteen of them crossed pairwise, leading to 73 pairs of heterokaryons having the same parental haploid nuclei, but mitochondria from either of the two parental isolates. In this study, we have used phenotypic data from these thirteen homokaryotic parents and their heterokaryotic offspring to investigate mitonuclear interactions.

Analysis of Phenotypic Data

Exploratory Data Analysis

A global analysis of data structure was carried out before comparing phenotypes of homokaryotic and heterokaryotic isolates. It shows that heterokaryons are characterized globally by their phenotypic homogeneity in comparison to their

homokaryotic parents, but can be divided in three groups: one small group of outliers with markedly lower maximal growth rate (3 isolates), a large group with high maximal growth rate and low phenotypic variance (107 isolates), and a medium group with high maximal growth rate, but a high phenotypic variance that seems intrinsic to their parental combinations (36 isolates).

Data Dispersion

Dispersion of growth rate data was analyzed both in the heterokaryons and their homokaryotic parents to detect those with high level of data dispersion, and, if so, to identify a plausible cause for it. Quartile coefficients of dispersion (QCD) of growth rate data were calculated at different levels of subdivision (within Petri dish, assay, or isolate). Their distributions show that differences of growth rate between assays (isolate level of data subdivision) contribute the most to data dispersion: while 0.76% of the dishes data and of 7.1% of the assays data have a QCD larger than 0.125, data of 38 heterokaryons (26 %) have a QCD larger than this value. A QCD larger than 0.125 was subsequently considered the hallmark of high level of data dispersion throughout this study. Data dispersion between assays in the parental homokaryons is weakly correlated to a lower median growth rate ($r^2 = 0.55714$). In the heterokaryons however, a high QCD is neither explained by a low median growth rate ($r^2 = 0.26644$), or by having one or two parents with a high QCD ($r^2 = 0.0155$). This analysis ruled out the possibility that high level of data dispersion could have experimental cause and rather points at specific heterokaryotic combinations to explain this phenomenon.

Data Distribution

Distribution of growth rate data was analyzed both in the heterokaryons and in their homokaryotic parents. Similar to what was previously published for the entire cohort of isolates (see **Figure 3** in Clergeot et al., 2019), data distribution of the thirteen parents is multimodal, indicating some genetic polymorphism for growth rate, whereas it is unimodal for the 146 heterokaryons, showing a homogenization of phenotypic variance in the offspring (**Supplementary Figure 1**). The center of the heterokaryon data distribution is shifted toward values recorded in the fastest parents only. Data of the assay with the highest median of each heterokaryon were then regrouped in a separate subset called high_med. Similarly, data of the assays with the lowest median were regrouped in another subset called low_med. Distributions of both subsets were compared. For high_med, distribution is narrow and symmetric, centered on a high value of growth rate (median = 7.25 mm/day), with very few outliers toward the lower values (**Figure 1A**). For low_med, distribution is broad and left-skewed, (**Figure 1B**), although its center remains close to high values of growth rate (median = 6.25 mm/day). Distribution of maximal growth rate (**Figure 2**) shows at the same time that almost all heterokaryons have the capacity to grow at rate only found in the fastest parents (**Figure 3**). Out of 146 maximal growth rate values, 143 are symmetrically distributed around a median very close to 8 mm/day and have an inter-quartile range (IQR) of 1.25 mm/day (**Figure 2**). Three outliers in low values of growth rate (max_out)

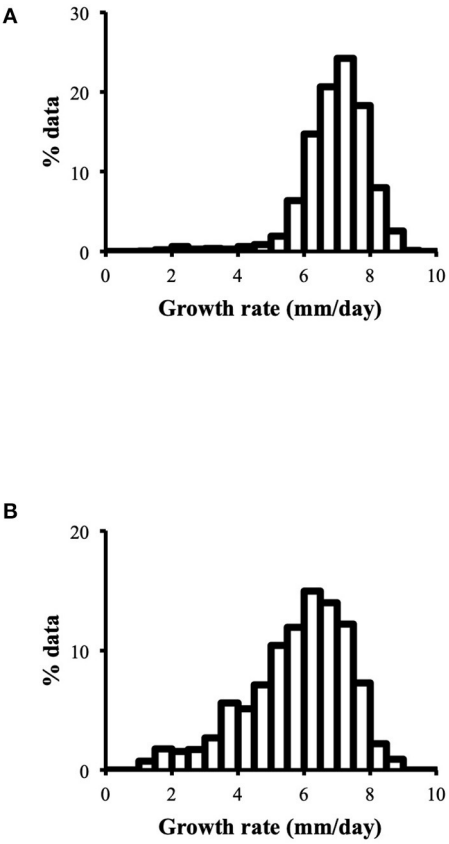


FIGURE 1 | (A) Data distribution of growth rate assays with the highest median of each heterokaryon is narrow and symmetric (high_med; $n = 1{,}563$); **(B)** Data distribution of growth rate assays with the lowest median of each heterokaryon is broad and left-skewed (low_med; $n = 1{,}487$); binwidth = 0.5.

correspond to heterokaryons 18(24), 24(18), and 9(32) and have maxima $\leq$ 4.75 mm/day. The 143 heterokaryons with high maxima (excluding the three outliers) were divided in two groups according to their QCDs: one group of 107 heterokaryons with a low data dispersion (QCD $\leq$ 0.125; subset named sta_het), another group of 36 heterokaryons with higher data dispersion (QCD $>$ 0.125; subset named var_het). Distributions of their maxima, third quartiles, medians, first quartiles, minima, and IQR were then compared (**Figure 4**). For a quarter of the heterokaryons, growth rate varies more between assays, but their maxima and third quartiles are similar to the ones of the 75% majority (**Figures 4A,B**). Spread and left-skewness of the data distribution of all heterokaryons is therefore rather explained by the larger phenotypic variance between assays recorded in these 36 isolates—shifting their median growth rate toward lower values (**Figures 4C–E**) and increasing their IQR (**Figure 4F**)—than by intrinsic difference of growth rate, observed for example between the slowest and the fastest parents (**Figure 3**).

Statistical Analyses of Phenotypic Data

Although the 143 heterokaryons with high maximal growth rate have similar genetic potential for this trait, their difference of

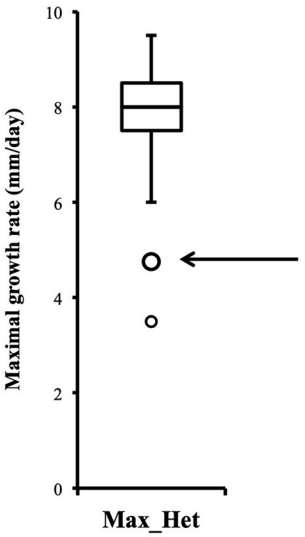


FIGURE 2 | Distribution of heterokaryon maximal growth rates is narrow, symmetric, and centered on a high value ($n = 146$). Maximal growth rate of heterokaryon 9(32) is one of the two outliers indicated by the arrow.

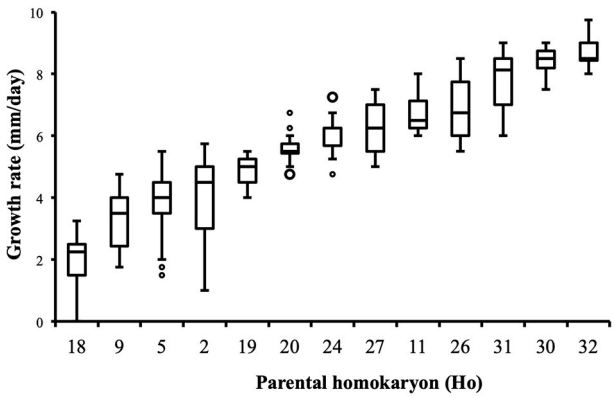


FIGURE 3 | Parental homokaryons display a large array of distributions of growth rate data. Maxima of the slowest parents are below minima of the fastest; $n = 24$ (Ho20, Ho24), $n = 48$ (Ho2, Ho5), or $n = 36$ (all other isolates).

phenotypic variance could bias statistical analyses based on their median growth rate. Therefore, two series of statistical analyses were performed with the 73 pairs of heterokaryons, first using all data, then using high_med data subset (complete results of the tests in **Supplementary Material 5**, tabs "MW tests—all data" and "MW tests—highest median"). Six pairs show a statistically significant difference of median growth rate in both series of tests (see **Table 1**). For heterokaryon pair (9;32), this support is the highest of all pairs using high_med. With this data subset, pair (9;32) has the largest difference of median growth rate of all pairs (3.5 mm/day), far above any other (next is (24;32) with 1,375 mm/day). Moreover, heterokaryon 9(32) displayed a low growth rate in two independent assays, something that only two other heterokaryons did, 18(24) and 24(18). Its maximal growth rate is a low outlier in the distribution of maxima of all heterokaryons (**Figure 2**). A complementary phenotypic analysis of pair (9;32) was carried out and compared with the one published in 2019 for the same pair (**Figure 5**). Repeated growth rate experiments with heterokaryons 9(32) and 32(9) confirmed their difference of phenotype [growth rate data from 2020 **Figure 5**; Mann-Whitney $U = 0$; n1 = n2 = 24; $\text{median}_{9(32)\_2} = 3{,}25$ mm/day; $\text{median}_{32(9)\_2} = 7{,}125$ mm/day; effect size $r = 0.8589$; $P < 0.001$; two tailed]. This data subset was used to show that 9(32) has indeed the slowest growth rate of all heterokaryons formed with Ho9 as acceptor parent, pointing out the fact that its combination with Ho32 is unique in this sample (**Figure 6**; complete results of statistical tests in **Supplementary Material 5**, tab "MW tests – Acc9").

Subsequent analysis focused on three specific pairs, (9;32), (18;32), and (9;18), for three reasons: (1) differences of phenotype of their heterokaryons are statistically grounded; (2) the complete offspring resulting from reciprocal crossings of their three parents is available for a grouped analysis; (3) based on results of previous SNP call and filtering, mitochondrial haplotypes of parents Ho9 and Ho18 were supposed to be identical, which was expected to facilitate genotyping. Using data of the assay with the highest median of each heterokaryon and each parent of these three pairs, a series of 36 statistical pairwise tests was carried out, in order to assess differences of growth rate (complete results of the tests in **Supplementary Material 5**, tab "MW tests – 3 pairs"). Difference of median growth rate is significant for all combinations, except 32(9)-32(18), 18(9)-18(32), and 9(32)-Ho9 (**Table 2**).

Genotypic Analysis

Phenotypic variance between the six isolates of pairs (9;32), (18;32), (9;18), and their three parents are best explained by the intervention of different alleles at a minimum of four nuclear loci, *nuc1*, *nuc2*, *nuc3*, and *nuc4*, and by non-additive variance due to three events of mitonuclear epistasis named mne1, mne2, and mne3 (**Figure 7**). Specific efforts were made to characterize *nuc2*, but also mne1 and mne2, as both presumably involve *nuc2*. The fact that heterokaryon 9(32) grows significantly slower than any other heterokaryon made with Ho9 as acceptor parent (**Figure 6**) shows that Ho9 and Ho32 are the only homokaryotic isolates to carry a recessive allele of *nuc2*. This allele, *nuc2-1*, interacts negatively with an allele of mitochondrial locus *mt2* borne by Ho9 only. Based on previous SNP calling and filtering, Ho9 and Ho18 were however supposed to have the same mitochondrial haplotype, which contradicts the fact that difference of growth rate between 9(18) and 18(9) is significant too. Other filtering of the SNPs called from nuclear and mitochondrial genome sequences of all homokaryotic isolates were set in order to identify markers linked to these two loci.

Mitochondrial Locus *mt2*

Filtering of mitochondrial SNPs was revisited and performed in a way that include indels. This led to the identification of three indels whose alleles are different in Ho9 and Ho18 (see complete list of 216 SNPs resulting from this filtering in **Supplementary Material 6**, tab "Mitochondrial SNPs"). Two

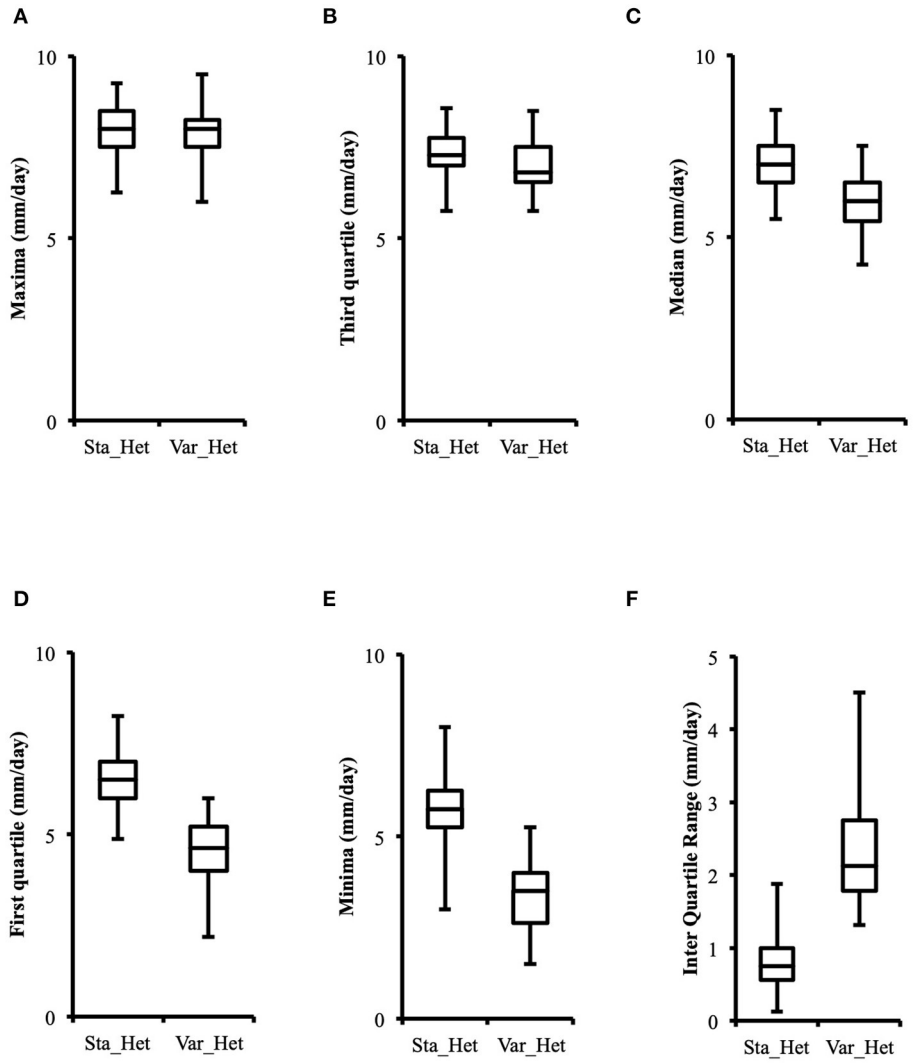


FIGURE 4 | Distributions of higher values of growth rate in phenotypically stable heterokaryons (sta_het; $n = 107$) and in phenotypically variable heterokaryons (var_het; $n = 36$) are similar, but the latter have a larger data spread toward lower values. **(A)** Maxima; **(B)** Third quartile; **(C)** Median; **(D)** First quartile; **(E)** Minima; **(F)** Inter quartile range.

were found in non-coding DNA, but one is localized in a non-conserved mitochondrial locus homologous to *nc-ORF5* in the mitochondrial genome sequence of *H. irregulare* (TAA/TAAA in position 472 of unitig_37; ORF is on the minus strand of this unitig). In *H. irregulare*, *nc-ORF5* is predicted to encode a protein of unknown function and made of 311 amino acid residues, annotated in the mitochondrial proteome of *H. irregulare* under NCBI accession YP009048485.1. Transcript and coding sequence of this locus were found to be shorter than predicted in the 5' end (181 aa instead of 311 in *H. irregulare*; see **Figure 8A** and magic-blast output in **Supplementary Material 7**), which is in agreement with the results of a protein similarity search showing that, although hypothetical orthologs of this protein exist in other Russulales fungi, no similar proteins are found in other species except for a hypothetical one in soilborne gammaproteobacteria with a relatively high sequence similarity and a similar length to nc-ORF5p (query length: 181 aa; subject length: 194 aa; query cover: 83%; 58% identity; E = 3e-51; subject GenBank ID: TLY78166.1). Consequently, it is reasonable to assume that this *H. parviporum* indel is not located within *nc-ORF5* itself, but either in its promoter, or its 5' untranslated transcript sequence (5' UTR). Interestingly, a second indel identified after mitochondrial SNP filtering is localized 44 bp downstream of the first SNP, in the promoter/5' UTR sequence of the same ORF (TAA/TAAA in position 428 of unitig_37; ORF is on the minus strand of this unitig). Ho9 and Ho18 have the same allele of this indel. Overall, three different alleles of *nc-ORF5* promoter/5' UTR exist in *H. parviporum* homokaryons used in this study, depending on the allelic combination of these two indels: Ho18 and Ho32 have a first allele *mt2-1*, Ho9, Ho11, Ho20, Ho26,

TABLE 1 | Pairs of heterokaryons with statistically different median growth rates; _het: all data of the heterokaryon were used; _high: data of the assay with the highest median of each heterokaryon were used; QCD: quartile coefficient of dispersion; Med: median growth rate; *p:* *p*-values of 2-tails Mann-Whitney tests; α: risk for type I error adjusted for multiple testing ($n = 73$; initial value = 0.05).

	Heterokaryon pair					
	(9;32)	**(30;32)**	**(2;11)**	**(9;18)**	**(18;32)**	**(24;32)**
het1	9(32)	30(32)	11(2)	9(18)	32(18)	32(24)
het2	32(9)	32(30)	2(11)	18(9)	18(32)	24(32)
QCD_het1	0.0698	0.1256	0.0890	0.0279	0.1415	0.0894
QCD_het2	0.2435	0.0625	0.0404	0.0660	0.2821	0.1351
Med_het1	4	6.25	6	5.5	6.5	7.5
Med_het2	5.625	7.75	6.75	6	5.5	5.875
p	0.00035	2.42E-06	5.27E-05	0.00076	7.67E-05	3.41E-08
α	0.00079	0.00069	0.00074	0.00081	0.00075	0.00068
QCD_high1	0.0588	0.0373	0.0400	0.0391	0.0407	0.0192
QCD_high2	0.0164	0.0186	0.0357	0.0467	0.0291	0.0446
Med_high1	4.25	7.625	6.125	5.625	7.75	8.125
Med_high2	7.75	8.25	6.75	6.625	6.5	6.75
p	3.03E-05	8.75E-05	0.00011	0.00013	0.00044	0.00075
α	0.00068	0.00069	0.00070	0.00071	0.00074	0.00076

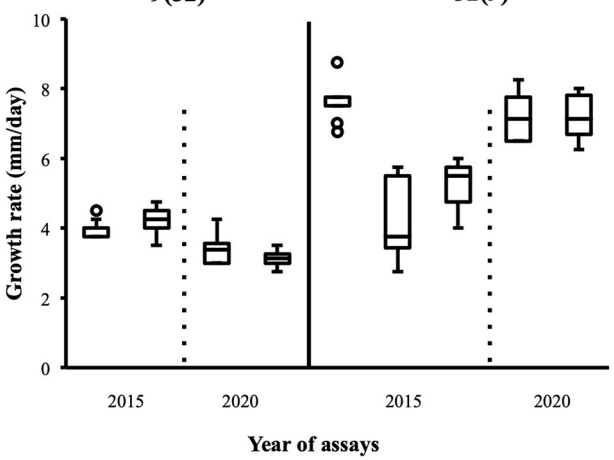


FIGURE 5 | Replicates of growth rate assays carried out with heterokaryon pair (9;32) confirm original phenotypes deduced from assays with the highest median; 2015: data published in Clergeot et al. (2019): data from this study ($n = 24$).

Ho27, Ho30, and Ho32 have a second allele *mt2-2*, and Ho2, Ho5, and Ho24 have a third one *mt2-3* (**Table 3**). In the specific context of the crossings of Ho9, Ho18, and Ho32, it is worth noticing that Ho18 and Ho32 have the same allele, *mt2-1*, while Ho9 has allele *mt2-2*.

The best way to explain the mitonuclear epistasis events observed (**Table 4**) is to invoke:

- for mne1, a negative interaction either between mitochondrial allele *mt2-2* and nuclear allele *nuc2-1* borne by Ho9 and Ho32 but not Ho18, or between mitochondrial allele *mt2-1* and nuclear allele *nuc2-2* borne by Ho18;
- for mne2, a partially negative interaction between mitochondrial allele *mt2-2* and nuclear allele *nuc2-2*, as observed in heterokaryon 9(18).

Nuclear Locus *nuc2*

From the phenotypes of all other heterokaryons obtained with parent Ho9 as nuclear acceptor, it was possible to deduce that homokaryotic isolates Ho2, Ho5, Ho11, Ho19, Ho20, Ho27, Ho26, Ho30, and Ho31 cannot bear recessive allele *nuc2-1* for sure (**Figure 6**). This information was used to set an allele-based filter for nuclear SNPs in order to provide candidate loci for *nuc2*, which led to the isolation of 1,242 SNPs. DNA of these SNP loci were first compared with a database made of all *H. irregulare* ORFs. For each locus matching an ORF, the SNP was localized in *H. parviporum* genomic DNA in order to identify those whose alternative allele causes a non-synonymous mutation. 449 SNPs were confirmed to be in an *H. parviporum* exon, of which 158 are responsible of a non-synonymous mutation in 78 different ORFs. Finally, the annotation of their *H. irregulare* orthologs were checked to select those associated with the mitochondria, either by localization or by function. This process of successive filtering led to the identification of two candidates for *nuc2* [see details in **Supplementary Material 6**, tab "Nuclear SNPs (non-syn)"]:

- A gene encoding a mitochondrial carrier protein (*H. irregulare* accession number: XM_009546414.1) with the following mutation events, all in exon 5 (**Figure 9A**): two consecutive nucleotide transversions (G to T and C to A) leading to an amino acid substitution G to V in position 303, the deletion of three consecutive nucleotides GCG leading to the deletion of amino acid A in position 309 in Ho7, Ho9, Ho12, Ho15, Ho32, and Ho35, and a nucleotide transversion G to A leading to amino-acid substitution R to K in position 330 in isolates Ho9 and Ho32;

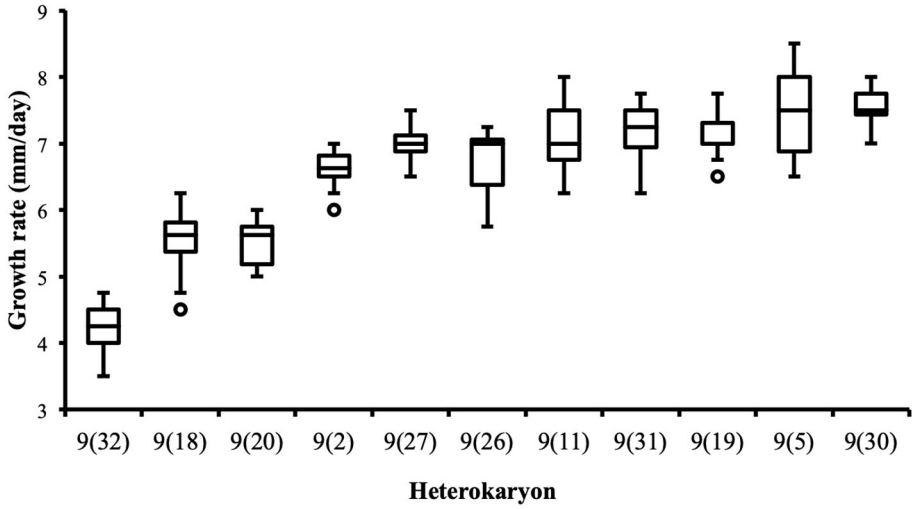


FIGURE 6 | Heterokaryon 9(32) grows significantly slower than any other heterokaryon made with Ho9 as acceptor parent. 2-tails Mann-Whitney tests were performed with data of the assays with the highest median for each isolate. Risk α for type I error was adjusted for multiple testing ($n = 10$; initial value = 0.05).

TABLE 2 | *p*-values of statistical tests performed to compare growth rates of heterokaryons from pairs (9;32), (9;18), and (18;32) and their parents Ho9, Ho18, and Ho32.

	9(32)	32(9)	18(32)	32(18)	9(18)	18(9)	Ho9	Ho18	Ho32
9(32)		3.03E-05	2.99E-05	3.12E-05	0.00015	3.08E-05	0.6821	2.88E-05	2.91E-05
32(9)	3.03E-05		0.00016	0.66174	3.05E-05	0.00029	2.97E-05	2.81E-05	9.70E-05
18(32)	2.99E-05	0.00016		0.00044	1.79E-04	0.64118	2.93E-05	2.78E-05	2.80E-05
32(18)	3.12E-05	0.66174	0.00044		3.15E-05	0.00107	3.06E-05	2.91E-05	2.93E-05
9(18)	0.00015	3.05E-05	0.00018	3.15E-05		0.00013	0.00013	2.91E-05	2.93E-05
18(9)	3.08E-05	0.00029	0.64118	0.00107	0.00013		3.01E-05	2.86E-05	2.88E-05
Ho9	0.68209	2.97E-05	2.93E-05	3.06E-05	0.00013	3.01E-05		2.83E-05	2.85E-05
Ho18	2.88E-05	2.81E-05	2.78E-05	2.91E-05	2.91E-05	2.86E-05	2.83E-05		2.70E-05
Ho32	2.91E-05	9.70E-05	2.80E-05	2.93E-05	2.93E-05	2.88E-05	2.85E-05	2.70E-05	

2-tails Mann-Whitney tests were performed with data of the assay with the highest median for each isolate; risk α for type I error was adjusted for multiple testing ($n = 36$; initial value = 0.05). For the combinations shaded in blue, the hypothesis of an identical data distribution could not be rejected.

- A gene encoding a hypothetical mitochondrial calcium uniporter (*H. irregulare* accession number: XM_009552795.1) with a single mutation due to a nucleotide transition C to G in exon 2 (**Figure 9B**) leading to amino acid substitution I to M in position 104 in Ho9, Ho23, Ho25, and Ho32.

Of the 793 SNPs not localized into an exon of a *H. parviporum* gene whose ortholog is transcribed in *H. irregulare*, 130 were found into an intron of such a gene, 86 were found <800 bp upstream of the first exon, and 68 <800 bp downstream of the last exon. Annotations of these genes were checked for possible association with the mitochondria, leading to the isolation of eight SNPs of interest [see details in **Supplementary Material 6**, tab "Nuclear SNPs (All)"]. Five SNPs were found into two introns of the gene encoding the mitochondrial carrier already mentioned (*H. irregulare* accession number: XM_009546414.1). One SNP was found 60 bp downstream of the stop codon of the gene encoding the hypothetical calcium uniporter already mentioned (*H. irregulare* accession number: XM_009552795.1). One SNP was found 113 bp upstream of the start codon of a gene encoding a beta subunit a of putative F1F0-type ATPase (*H. irregulare* accession number: XM_009544597.1). One SNP was found 11 bp upstream of the start codon of a gene encoding another mitochondrial carrier protein (*H. irregulare* accession number: XM_009545544.1).

DISCUSSION

Our phenotypic data show that two major competing trends are at work in heterokaryons, with opposite effects on growth rate. The first trend is "heterokaryon vigor" (Clergeot et al., 2019), which results from the reciprocal masking of deleterious alleles brought by one or both parents (Clark and Anderson, 2004). Our growth rate data show that dominance is frequent, but overdominance is rare and limited to the offspring of parents that are both less fit. As the fittest parental alleles are overwhelmingly dominant, the center of the distribution of all

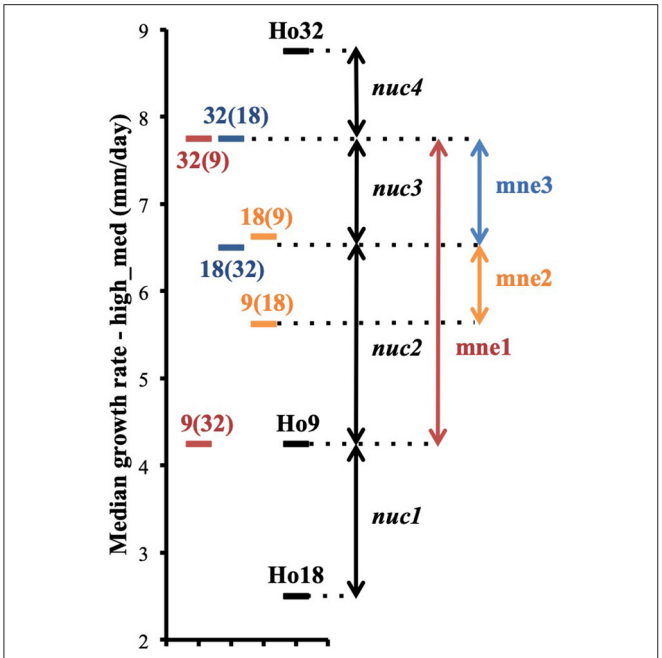


FIGURE 7 | Phenotypic differences between heterokaryons from pairs (9;32), (9;18), (18;32) and their parents are explained by the contribution of four different nuclear genetic loci *nuc1*, *nuc2*, *nuc3*, and *nuc4* and the intervention of three events of mitonuclear epistasis mne1, mne2, and mne3 (phenotypes are based on data of the assay with the highest median for each isolate; median growth rates are plotted on the figure).

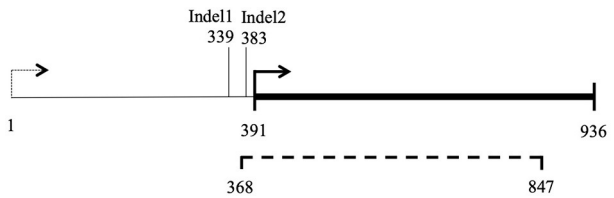


FIGURE 8 | Map of locus *nc-ORF5*. 1–936 bp: ORF prediction in *H. irregulare* mitochondrial genome annotation; Dashed line (368–847 bp): transcription detected in *H. irregulare* isolate TC32-1; Thick line (391–936 bp): actual *nc-ORF5*.

heterokaryons data is shifted toward the position it would have in a hypothetic homokaryon having the largest combination of the fittest parental alleles. Homokaryons growth rate data suggest that the fastest parents in our sample are close to such an optimal allelic combination, as none of the heterokaryotic offspring can grow at a higher rate than they do.

The second trend is high heterokaryon-specific phenotypic variation, a phenomenon that we have observed in several heterokaryons (26%) after carrying out different growth rate assays over time. These heterokaryons can nevertheless grow as fast as the others. Their genetic potential for this specific trait is therefore not different from the rest of the heterokaryons. The fact that their phenotype does vary more between assays is due to other factors, unidentified so far, although probably linked to

TABLE 3 | Mitochondrial locus *mt2* exists in three different forms, *mt2-1*, *mt2-2*, and *mt2-3*, in the thirteen *H. parviporum* parental homokaryons depending on their indels alleles.

Homokaryotic isolate (Ho)	Indel1 allele	Indel2 allele	Locus allele
18; 32	TAA	TAA	*mt2-1*
9; 11; 19; 20; 26; 27; 30; 31	TAAA	TAA	*mt2-2*
2; 5; 24	TAAA	TAAA	*mt2-3*

the heterokaryotic state, during which parental haploid nuclei coexist in the same cytoplasm, but do not fuse (Raper, 1966). The heterokaryon is an intermediary state between haploidy and true diploidy, more plastic than the latter due to the possibilities of remating (Hansen et al., 1993; Johannesson and Stenlid, 2004; James et al., 2009), secondary formation of haploid sectors (Stenlid and Rayner, 1991; Hansen et al., 1993), and production of haploid conidia (Hansen et al., 1993). Although considered as functionally equivalent to a diploid (Day and Roberts, 1969), such a dynamic and possibly conflicting association of haploid nuclei is likely to lead to stochastic variation of gene expression (Kilfoil et al., 2009), and therefore, of heterokaryon phenotype. Degree of phenotypic variation of each heterokaryon might be genetically determined by specific combination of parental alleles, as for nuclear ratio in heterokaryons of *H. parviporum* (James et al., 2008).

Analysis of heterokaryons growth rate data showed that higher values reflect their genetic potential for this trait better, which made possible the identification of pairs of heterokaryons issued from the same parents, but whose phenotypes are different. It confirms our previous detection of mitochondrial genetic effects in the same cohort of heterokaryons of *H. parviporum* (Clergeot et al., 2019). In one occurrence, owing to the facts that two non-kin-related parental isolates (Ho9 and Ho18) have nearly identical mitochondrial haplotypes and that their crossings with a third isolate, Ho32, gave heterokaryons with clear-cut phenotypes, we have identified alleles of a mitochondrial locus (*mt2*) and a genomic locus (*nuc2*) involved in a genetic interaction in *trans* across the two genomes. Different combination of alleles at these two loci contribute to phenotypic variation in the offspring of the crossings with these three isolates.

Candidates for genomic locus *nuc2* were searched using phenotypic information of heterokaryons from all other crossings involving parent Ho9 and SNPs markers called from the genome sequences of all homokaryotic isolates (Clergeot et al., 2019). In this work, SNP alleles associated with a non-synonymous mutation in two *H. parviporum* ORFs were investigated in priority as more likely candidates for *nuc2*, although other causes of phenotypic variation cannot be excluded at this stage. Similar bioinformatics methods have been used to track polymorphism in gene regulatory sequences or introns too, and led to the identification of two additional candidates for *nuc2*. The two genes whose sequences in Ho9 and Ho32 are bearing mutations altering the structure of their translation products (although it cannot be ascertained that the function of these products is modified), but also carrying mutations that might affect their expression level are considered the best possible

TABLE 4 | Expression of nuclear locus *nuc2* depends on appropriate allelic combination with mitochondrial locus *mt2*; E: full expression of the trait; e: partial expression; –: no expression; *nuc2* (d): *nuc2* allele brought by the donor isolate in heterokaryons.

	nuc2-1	***nuc2-2***							
mt2-1	E	-							
mt2-2	–	e							
	Ho9	**Ho18**	**Ho32**	**9(18)**	**18(9)**	**9(32)**	**32(9)**	**18(32)**	**32(18)**
nuc2	*nuc2-1*	*nuc2-2*	*nuc2-1*	*nuc2-1*	*nuc2-2*	*nuc2-1*	*nuc2-1*	*nuc2-2*	*nuc2-1*
nuc2 (d)				*nuc2-2*	*nuc2-1*	*nuc2-1*	*nuc2-1*	*nuc2-1*	*nuc2-2*
mt2	*mt2-2*	*mt2-1*	*mt2-1*	*mt2-2*	*mt2-1*	*mt2-2*	*mt2-1*	*mt2-1*	*mt2-1*
Expression	–	–	E	e	E	–	E	E	E

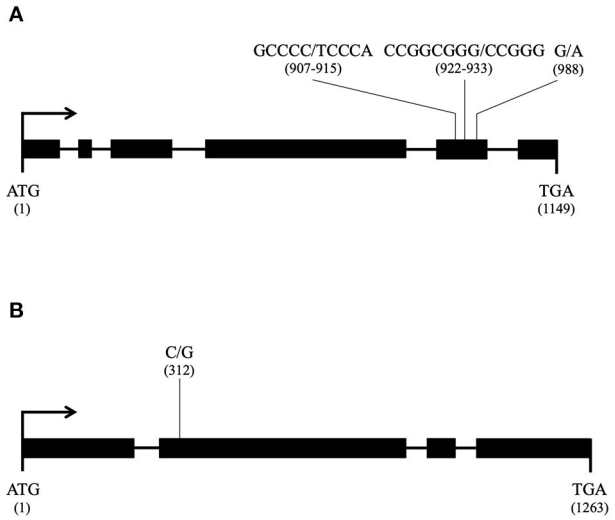


FIGURE 9 | Candidate nuclear loci for *nuc2* in *Heterobasidion parviporum* and their mutation profile (thick line: exon; thin line: intron; for each mutation, the reference allele is given first and its position in the coding sequence is in parentheses); **(A)** Gene encoding a mitochondrial carrier protein orthologous to *Heterobasidion irregulare* XM_009546414.1; **(B)** Gene encoding a mitochondrial calcium uniporter orthologous to *Heterobasidion irregulare* XM_009552795.1.

candidates at this stage. Both are involved in molecular trafficking to the mitochondrion (a carrier protein and a calcium uniporter). Although the gene encoding a calcium uniporter cannot be ruled out at this stage, the other one encoding a mitochondrial carrier protein is a more likely candidate for *nuc2* for several reasons: Mutations in this gene result from multiple transversions (3) and one indel in the same exon, whereas the mutation in the gene encoding the calcium uniporter results from a single transition; A multiple protein alignment with orthologs from other species shows that substitution R to K in the sequence of the carrier is in a conserved motif, and this substitution is present in Ho9 and Ho32 only; Amino-acid substitutions/deletion observed in the carrier are more likely to lead to a change of balance between structure and function.

In contrast, it is not a mitochondrial gene encoding a key protein in oxidative metabolism that has been uncovered as the potential candidate involved in a mitonuclear interaction with *nuc2*, but a non-conserved gene instead, with orthologs in the Russulales order only. The ortholog of this gene in the mitochondrial genome of *H. irregulare* has been previously identified as *nc-ORF5* (non-conserved Open Reading Frame 5; Himmelstrand et al., 2014). The presence of nc-genes in mitochondrial DNA is a typical feature of Agaricomycetes fungi. Nc-genes are supposed to share the same evolutionary origin as the mitochondrial plasmid DNA frequently found in the mitochondrial genome of species in this class of fungi and contribute to explain its unusually large size. This non-conserved and exchangeable part of Agaricomycetes mitochondrial DNA might originate from the uptake of plasmids by mitochondria and its subsequent integration into mitochondrial DNA early in the evolution of this class of fungi (Himmelstrand et al., 2014; Medina et al., 2020). Protein blast searches made with the translation of the ortholog of *nc-ORF5* in *H. parviporum* (Hparv-nc-ORF5p) and by excluding proteins from the Russulales retrieved exclusively a hypothetical protein from soilborne gammaproteobacteria (Diamond et al., 2019). Phenotypic variation associated with mitochondrial DNA sequence polyporphism at this locus seems to originate from a change in the regulation of *nc-ORF5* expression rather than from its loss of function, as the SNP is located upstream of *nc-ORF5* coding sequence, either in its promoter or its 5' UTR sequence.

Further investigation would imply (1) studying the expression of all mitochondrial and nuclear candidate genes to confirm or infirm hypotheses made about them, (2) providing molecular genetics evidence of mitonuclear interaction by carrying complementation experiments of homokaryon Ho9 with wild-type copies of candidate nuclear loci for *nuc2*. Genetic transformation of *Heterobasidion spp*. using *Agrobacterium tumefaciens* has been described previously (Samils et al., 2006), including in *H. parviporum* (Ihrmark, *unpubl.*), and could be used for that purpose. Another option would be to shuffle mitochondria into various homokaryotic nuclear backgrounds by reisolating homokaryotic conidia from heterokaryons cultures (Ramsdale and Rayner, 1994) and assessing their phenotype. Finally, although our phenotypic analysis points at a very simple genetic explanation for the phenotypic variation observed in this sample of isolates (homogeneization of the heterokaryons phenotypes to the level observed in the best homokaryotic

parents is the signature of the complementation of a few recessive and defective alleles unique to each slower parent by dominant wild-type alleles borne by any other parent), mitonuclear interactions involving more than one locus in the mitochondrial and/or the nuclear genome cannot be excluded so far. It could be determined by carrying out a segregation analysis (Lind et al., 2005).

Mitonuclear interactions are suspected to be the cause of some breakthrough increase of virulence or change of host range in rare interspecific hybrids (Olson and Stenlid, 2001; Giordano et al., 2018; Hu et al., 2020), but they are also part of the ordinary evolutionary process of each fungal pathogenic species and could hence contribute to mold the interactions with its host. Coadaptation between mitochondrial and nuclear DNA might be an important driver in the evolution of host-pathogen interactions. Several examples have been documented in other fungal pathogens of a change of host range or of a variation in virulence associated either with polymorphism of mitochondrial DNA, with a loss of function, or with a differential expression of nuclear genes encoding mitochondrial proteins (Monteiro-Vitorello et al., 1995; Lorenz and Fink, 2001; Inoue et al., 2002; Zhan et al., 2004; Ma et al., 2009; Mahlert et al., 2009; Patkar et al., 2012; Zhang et al., 2012; Khan et al., 2015; van de Vossenberg et al., 2018). In the context of this work, it is worth noticing that a nuclear-encoded mitochondrial carrier protein has already been mentioned in *Heterobasidion* spp. as possibly involved in virulence variation related to mitochondrial background (Giordano et al., 2018; Hu et al., 2020). Furthermore, in pathogenic Basidiomycetes fungi like *Heterobasidion spp.*, the capacity of heterokaryon remating (Johannesson and Stenlid, 2004) gives mitonuclear interactions the opportunity to add to the selection process for fitter heterokaryotic associations while genetically diverse heterokaryons grow, meet and exchange nuclei on host wood. Our work brings forth evidence that the investigation of mitonuclear interactions is possible for a trait like hyphal growth rate, but the same could be considered for virulence.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

P-HC and ÅO: conceived and designed the experiments, performed the experiments, analyzed the data, and revised the paper. P-HC: exploratory data analysis, statistics, and bioinformatics and wrote the paper. ÅO: additional growth rate assays. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Carl Tryggers Foundation for Scientific Research.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Frederik Witte (Universistets- och Högskolerådet, Sweden) for comments and suggestions about data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/ffunb.2021.779337/full#supplementary-material

REFERENCES

Allio, R., Donega, S., Galtier, N., and Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* 34, 2762–2772. doi: 10.1093/molbev/msx197

Bernhardt, D., Hamann, A., and Osiewacz, H. D. (2014). The role of mitochondria in fungal aging. *Curr. Opin. Microbiol.* 22, 1–7. doi: 10.1016/j.mib.2014.09.007

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, P709–P716. doi: 10.1016/j.tig.2003.10.012

Clark, T. A., and Anderson, J. B. (2004). Dikaryons of the Basidiomycete fungus *Schizophyllum commune*: evolution in long-term culture. *Genetics* 167, 1663–1675. doi: 10.1534/genetics.104.027235

Clergeot, P.-H., Rode, N. O., Glémin, S., Brandström-Durling, M., Ihrmark, K., and Olson, Å. (2019). Estimating the fitness effect of deleterious mutations during the two phases of the life cycle: a new method applied to the root-rot fungus *Heterobasidion parviporum*. *Genetics* 211, 963–976. doi: 10.1534/genetics.118.301855

Day, P. R., and Roberts, C. F. (1969). Complementation in dikaryons and diploids of *Coprinus lagopus*. *Genetics* 62, 265–270. doi: 10.1093/genetics/62.2.265

Diamond, S., Andeer, P. F., Li, Z., Crits-Christoph, A., Burstein, D., Anantharaman, K., et al. (2019). Mediterranean grassland soil C-N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat. Microbiol.* 4, 1356–1367. doi: 10.1038/s41564-019-0449-y

Garbelotto, M., and Gonthier, P. (2013). Biology, epidemiology, and control of *Heterobasidion* species worldwide. *Annu. Rev. Phytopathol.* 51, 39–59. doi: 10.1146/annurev-phyto-082712-102225

Giordano, L., Sillo, F., Garbelotto, M., and Gonthier, P. (2018). Mitonuclear interactions may contribute to fitness of fungal hybrids. *Sci. Rep.* 8:1706. doi: 10.1038/s41598-018-19922-w

Gray, M. W., Burger, G., and Lang, B. F. (1999). Mitochondrial evolution. *Science* 283, 1476–1481. doi: 10.1126/science.283.5407.1476

Gray, M. W., Burger, G., and Lang, B. F. (2001). The origin and early evolution of mitochondria. *Genome Biol.* 2, 1018.1–1018.5. doi: 10.1186/gb-2001-2-6-reviews1018

Hansen, E. M., Stenlid, J., and Johansson, M. (1993). Somatic incompatibility and nuclear reassortment in *Heterobasidion annosum*. *Mycol. Res.* 97, 1223–1228. doi: 10.1016/S0953-7562(09)81289-6

Henze, K., and Martin, W. (2003). Essence of mitochondria. *Nature* 426, 127–128. doi: 10.1038/426127a

Himmelstrand, K., Olson, Å., Brandström-Durling, M., Karlsson, M., and Stenlid, J. (2014). Intronic and plasmid-derived regions contribute to the large mitochondrial genome sizes of Agaricomycetes. *Curr. Genet.* 60, 303–313. doi: 10.1007/s00294-014-0436-z

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Holt, C. E., Gockel, H., and Hüttermann, A. (1983). The mating system of *Fomes annosus*. *Eur. J. Forest Pathol.* 13, 174–181. doi: 10.1111/j.1439-0329.1983.tb01415.x

Hu, Y., Elfstrand, M., Stenlid, J., Brandström-Durling, M., and Olson, Å. (2020). The conifer root rot pathogens *Heterobasidion irregulare* and *Heterobasidion occidentale* employ different strategies to infect Norway spruce. *Sci Rep*. 10:5884. doi: 10.1038/s41598-020-62521-x

Inoue, I., Namiki, F., and Tsuge, T. (2002). Plant colonization by the vascular wilt fungus *Fusarium oxysporum* requires *FOW1*, a gene encoding a mitochondrial protein. *Plant Cell* 14, 1869–1883. doi: 10.1105/tpc.002576

James, T. Y., Johansson, S. B. K., and Johannesson, H. (2009). Trikaryon formation and nuclear selection in pairings between heterokaryons and homokaryons of the root rot pathogen *Heterobasidion parviporum*. *Mycol. Res*. 113, 583–590. doi: 10.1016/j.mycres.2009.01.006

James, T. Y., Stenlid, J., Olson, Å., and Johannesson, H. (2008). Evolutionary significance of imbalanced nuclear ratios within heterokaryons of the basidiomycete fungus *Heterobasidion parviporum*. *Evolution* 62, 2279–2296. doi: 10.1111/j.1558-5646.2008.00462.x

Johannesson, H., and Stenlid, J. (2004). Nuclear reassortment between vegetative mycelia in natural populations of the basidiomycete *Heterobasidion annosum*. *Fungal Genet. Biol*. 41, 563–570. doi: 10.1016/j.fgb.2004.01.002

Khan, I. A., Ning, G., Liu, X., Feng, X., Lin, F., and Lu, J. (2015). Mitochondrial fission protein MoFis1 mediates conidiation and is required for full virulence of the rice blast fungus *Magnaporthe oryzae*. *Microbiol. Res*. 178, 51–58. doi: 10.1016/j.micres.2015.06.002

Kilfoil, M. L., Lasko, P., and Abouheif, E. (2009). Stochastic variation: from single cells to superorganisms. *HFSP J*. 3, 379–385. doi: 10.2976/1.3223356

Kurland, C. G., and Andersson, S. G. E. (2000). Origin and evolution of the mitochondrial genome. *Microbiol. Mol. Biol. Rev*. 64, 786–820. doi: 10.1128/MMBR.64.4.786-820.2000

Lind, M., Olson, Å., and Stenlid, J. (2005). An AFLP-marker based genetic linkage map of *Heterobasidion annosum* locating intersterility genes. *Fungal Genet. Biol*. 42, 519–527. doi: 10.1016/j.fgb.2005.03.005

Lorenz, M. C., and Fink, G. R. (2001). The glyoxylate cycle is required for fungal virulence. *Nature* 412, 83–86. doi: 10.1038/35083594

Ma, H., Hagen, F., Stekel, D. J., Johnston, S. A., Sionov, E., Falk, R., et al. (2009). The fatal fungal outbreak on Vancouver Island is characterized by enhanced intracellular parasitism driven by mitochondrial regulation. *Proc. Natl. Acad. Sci. U.S.A*. 106, 12980–12985. doi: 10.1073/pnas.0902963106

Mahlert, M., Vogler, C., Stelter, K., Hause, G., and Basse, C. W. (2009). The *a2* mating-type-locus gene *lga2* of *Ustilago maydis* interferes with mitochondrial dynamics and fusion, partially in dependence on a Dnm1-like fission component. *J. Cell. Sci*. 122, 2402–2412. doi: 10.1242/jcs.039354

McBride, H. M., Neuspiel, M., and Wasiak, S. (2006). Mitochondria: More than just a powerhouse. *Curr. Biol*. 16, R551–R560. doi: 10.1016/j.cub.2006.06.054

Medina, R., Franco, M. E. E., Bartel, L. C., Martinez, A. V., Saparrat, M. C. N., and Balatti, P. A. (2020). Fungal mitogenomes: relevant features to planning plant disease management. *Front. Microbiol*. 11:978. doi: 10.3389/fmicb.2020.00978

Monteiro-Vitorello, C. B., Bell, J. A., Fulbright, D. W., and Bertrand, H. (1995). A cytoplasmically transmissible hypovirulence phenotype associated with mitochondrial DNA mutations in the chestnut blight fungus *Cryphonectria parasitica*. *Proc. Natl. Acad. Sci. U.S.A*. 92, 5935–5939. doi: 10.1073/pnas.92.13.5935

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. doi: 10.1186/s12862-020-01685-6

Nieuwenhuis, B. P., Billiard, S., Vuilleumier, S., Petit, E., Hood, M. E., and Giraud, T. (2013). Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity* 111, 445–455. doi: 10.1038/hdy.2013.67

Olson, Å., Aerts, A., Asiegbu, F., Belbahri, L., Bouzid, O., Broberg, A., et al. (2012). Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytol*. 194, 1001–1013. doi: 10.1111/j.1469-8137.2012.04128.x

Olson, Å., and Stenlid, J. (2001). Mitochondrial control of fungal hybrid virulence. *Nature* 411:438. doi: 10.1038/35078147

Patkar, R. N., Ramos-Pamplona, M., Gupta, A. P., Fan, Y., and Naqvi, N. I. (2012). Mitochondrial β-oxidation regulates organellar integrity and is necessary for conidial germination and invasive growth in *Magnaporthe oryzae*. *Mol. Microbiol*. 86, 1345–1363. doi: 10.1111/mmi.12060

Ramsdale, M., and Rayner, A. D. M. (1994). Distribution patterns of number of nuclei in conidia from heterokaryons of *Heterobasidion annosum* (Fr.) Bref. and their interpretation in terms of genomic conflict. *New Phytol*. 128, 123–134. doi: 10.1111/j.1469-8137.1994.tb03995.x

Raper, J. R. (1966). *Genetics of Sexuality in Higher Fungi*. New York, NT: Ronald Press. doi: 10.1126/science.151.3708.315

Samils, N., Elfstrand, M., Lindner Czederpiltz, D. L., Fahleson, J., Olson, Å., Dixelius, C., et al. (2006). Development of a rapid and simple *Agrobacterium tumefaciens*-mediated trasnformation system for the fungal pathogen *Heterobasidion annosum*. *FEMS Microbiol. Lett*. 205, 82–88. doi: 10.1111/j.1574-6968.2005.00069.x

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

Stenlid, J., and Rayner, A. D. M. (1991). Patterns of nuclear migration and heterokaryosis in pairings between sibling homokaryons of *Heterobasidion annosum*. *Mycol. Res*. 95, 1275–1283 doi: 10.1016/S0953-7562(09)80574-1

Stojković, B., and Dordević, M. (2017). Interaction between mitochondrial and nuclear genomes: the role in life-history evolution. *Biol. Serb*. 39, 32–40. doi: 10.5281/zenodo.826619

Urbani, A., and Babu, M. (2019). Mitochondria in health and sickness. *Adv. Exp. Med. Biol*. 1158, 119–142. doi: 10.1007/978-981-13-8367-0

van de Vossenberg, B. T. L. H., Brankovics, B., Nguyen, H. D., 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

Xu, J., and Wang, P. (2015). Mitochondrial inheritance in basidiomycete fungi. *Fungal Biol. Rev*. 29, 209–219. doi: 10.1016/j.fbr.2015.02.001

Zhan, J., Kema, G. H., and McDonald, B. A. (2004). Evidence for natural selection in the mitochondrial genome of *Mycosphaerella graminicola*. *Phytopathology* 94, 261–267. doi: 10.1094/PHYTO.2004.94.3.261

Zhang, X.-W., Jia, L.-J., Zhang, Y., Jiang, G., Li, X., Zhang, D., et al. (2012). In planta stage-specific fungal gene profiling elucidates the molecular strategies of *Fusarium graminearum* growing inside wheat coleoptiles. *Plant Cell* 24, 5159–5176. doi: 10.1105/tpc.112.105957

Conflict of Interest: The 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.

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.

frontiers in Microbiology

MINI REVIEW
published: 20 January 2022
doi: 10.3389/fmicb.2022.806575

From Genome Variation to Molecular Mechanisms: What we Have Learned From Yeast Mitochondrial Genomes?

*Weilong Hao**

Department of Biological Sciences, Wayne State University, Detroit, MI, United States

Analysis of genome variation provides insights into mechanisms in genome evolution. This is increasingly appreciated with the rapid growth of genomic data. Mitochondrial genomes (mitogenomes) are well known to vary substantially in many genomic aspects, such as genome size, sequence context, nucleotide base composition and substitution rate. Such substantial variation makes mitogenomes an excellent model system to study the mechanisms dictating mitogenome variation. Recent sequencing efforts have not only covered a rich number of yeast species but also generated genomes from abundant strains within the same species. The rich yeast genomic data have enabled detailed investigation from genome variation into molecular mechanisms in genome evolution. This mini-review highlights some recent progresses in yeast mitogenome studies.

Keywords: mutation, recombination, repeat, mobile introns, gene conversion, GC-content

OPEN ACCESS

Edited by:
Georg Hausner,
University of Manitoba, Canada

Reviewed by:
Heather Fiumera,
Binghamton University, United States
Feng Ling,
RIKEN, Japan
Hsin-Han Lee,
Academia Sinica, Taiwan

***Correspondence:**
Weilong Hao
haow@wayne.edu

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *01 November 2021*
Accepted: *03 January 2022*
Published: *20 January 2022*

Citation:
Hao W (2022) From Genome Variation to Molecular Mechanisms: What we Have Learned From Yeast Mitochondrial Genomes? Front. Microbiol. 13:806575. doi: 10.3389/fmicb.2022.806575

INTRONDUCTION

Mitochondrial genomes (mitogenomes) originated from an alpha-proteobacterium *via* endosymbiosis (Lang et al., 1999), and have adopted radically different shapes, sizes, and organizations (Burger et al., 2003; Shao et al., 2009; Sloan et al., 2012; Smith et al., 2012). The great variation of mitogenome diversity and complexity has revolutionized our view of genome evolution and facilitated development of new evolutionary theories (Bazin et al., 2006; Lynch et al., 2006; Whitney and Garland, 2010; Sloan et al., 2012; Christensen, 2013). The excitement of mitogenomes inspired more sequencing projects, and perhaps more importantly, many mindful and in-depth comparative genomics studies (Smith, 2015). These efforts continue to push the boundaries of our understanding in genome evolution.

Mitogenomes are highly variable among yeast species. Mitogenomes show substantial differences in genome size and organization, GC-content, mutation rates, and recombination frequencies. As mutation is the fundamental source of genetic variation, mitogenome differences provide important insights into the underlying mechanisms in the mutation processes. The budding yeast *Saccharomyces cerevisiae* is among the best-studied model organisms, an abundant number of mitogenomes have been sequenced for *S. cerevisiae* (Strope et al., 2015; De Chiara et al., 2020) and related species (Freel et al., 2014; Nguyen et al., 2020b). The abundant genomic data have allowed comparative analyses of mitogenomes among closely related yeast species and also intraspecific strains within a species. Detailed variations among closely related mitogenomes become uniquely informative to help identify molecular mechanisms driving the genomic changes.

MITOCHONDRIAL DNA DELETION

The mitochondrial DNAs (mtDNAs) of related yeast strains is well known to show hypervariability in sequence [reviewed in Borst and Grivell (1978)]. Studies of hypervariable yeast mitogenomes

have led to discoveries of new molecular and cellular mechanisms (Dujon, 2020). Yeast spontaneously loses mtDNAs and develops respiratory-incompetent petite colonies (rho$^-$) (Whittaker et al., 1972). Petite colonies arise naturally under normal growth conditions, but the frequency of petite formation varies among strains. In *S. cerevisiae*, the frequency of developing spontaneous petite colonies varies by about 100-fold between laboratory strains versus natural isolates (Dimitrov et al., 2009). The formation of petite colonies, or petite-positive trait, has been associated with whole-genome duplication of the nuclear genome (Piskur, 2001), which provides the basis for neofunctionalization. A survey in more than one hundred strains, however, shows that the petite positive trait is throughout the Saccharomycetaceae family, much beyond the whole genome duplication species (Fekete et al., 2007). Some deletions in yeast mtDNAs are mediated by gene conversion between GC cluster repeats (Weiller et al., 1991), which are characterized by high GC content and palindromic structure (Yin et al., 1981) (also described below). Petite formation, as a genetic trait, has been studied in genome-wide association studies to identify the associated genes (Dimitrov et al., 2009). The different frequencies in petite formation among strains have been associated with genetic variation in at least four nuclear-encoded genes, the mtDNA polymerase (MIP1) (Baruffini et al., 2007a) and other less-well studied SAL1, CAT5, and MKT1 genes (Dimitrov et al., 2009).

MOVEMENT OF MITOCHONDRIAL INTRONS

Genetic crosses of different mitochondrial genotypes have been conducted in *S. cerevisiae* to understand yeast mitochondrial inheritance and genetics (Wilkie and Thomas, 1973). Mitochondrial markers are inherited in a non-mendelian manner (Jacquier and Dujon, 1985). A well-studied example is the omega (ω) locus in 21S rRNA (Butow, 1985), in which the ω+ strains can transfer the whole intron unidirectionally to an intron-less locus (Dujon et al., 1974). The ω locus belongs to group I intron, a class of self-catalytic ribozymes that often encode a homing endonuclease gene (HEG) (Perlman and Butow, 1989). The self-spicing intron and intron-encoded endonuclease play an important role in driving intron invasion and mobility (Dujon, 1989). The mobility and invasion of the ω intron can go beyond the species boundary. The phylogeny of ω-HEG is significantly different from that of the host 21S rRNA (Goddard and Burt, 1999), suggesting cross-species horizontal transfer of ω-HEG. Given the highly invasive nature of the ω intron, one would expect most yeast strains to be ω+. Yet, many yeast strains remain ω–. Goddard and Burt (1999) provided a framework to explain the sporadic distribution of the ω intron; after invasion, the omega-HEGs undergo rapid degeneration and loss, followed by reinvasion. This process is coined as the Goddard–Burt life cycle by Mukhopadhyay and Hausner (2021). The actual life cycle of introns is likely much more complex than cyclical invasion, degeneration, loss and then reinvasion.

In the Saccharomycetaceae family, mitochondrial introns are found in three genes, *cox1*, *cob,* and 21S rRNA, in total 17 intron positions (Wu et al., 2015). Each intron has a unique distribution pattern, and intron content often varies substantially even among small numbers of conspecific strains within species. Except the *cox1* i1 intron, all other 16 introns are sporadically distributed. The evolutionary turnover rates of gain and loss among these introns were quantitatively measured (Wu et al., 2015). The high-mobility introns documented in genetic crosses do not necessarily have faster turnover rates than low-mobility introns. The *cox1* i1 intron is currently only found in *S. cerevisiae* strains, with a high rate of intron insertion, it will not be surprising if the *cox1* i1 intron is present in some upcoming non-*S. cerevisiae* mitogenomes. Furthermore, phylogenetically mosaic sequences are evident in both introns and HEGs (Wu and Hao, 2014). Thus, intron and its encoded HEG do not always transmit together as a unit. These findings support that gene conversion between the donor and recipient sequences can take place at both the gene and intragenic levels and lead to insertion or deletion of the adjacent HEG/intron content (Wu and Hao, 2014).

ELEVATED SEQUENCE EVOLUTION NEAR MOBILE INTRONS

The distribution of mutation along DNA sequences is not uniform, and yeast mitogenomes are no exception. The recent availability of abundant yeast population genomic data makes it possible to examine genetic diversity along mitochondrial genes. One striking finding is the increased density of single nucleotide polymorphisms (SNP) in exon regions approaching intron boundaries (Repar and Warnecke, 2017). Although intron mobility is recognized to play a critical role in driving the sequence diversity of host genes, the underlying mechanisms cannot be easily identified. There are two possible mechanisms that can increase SNP density in exons (Repar and Warnecke, 2017). First, diverse exonic sequences are gene conversion tracts, or known as co-conversion tracts, acquired from distantly related species. Horizontal transfer of introns and co-conversion tracts has been well documented in plant mitogenomes (Cho and Palmer, 1999). Since most plant mitogenomes have exceptionally low mutation rates, shared long co-conversion tracts among distantly related intron-containing sequence can be easily and convincingly identified (Sanchez-Puerta et al., 2011). Similar to plant mitochondrial introns, yeast mitochondrial introns also undergo horizontal transfer (Wu et al., 2015) and gene conversion at intragenic level (Wu and Hao, 2014). The relative high sequence divergence among yeast mitogenomes makes it challenging to accurately identify the donor species of the diverse co-conversion tracts. Alternatively, the flanking regions of each intron insert site are mutation hotspots because of endonuclease activity and subsequent error-prone repair. Yeast mitogenomes contain stand-alone HEGs, which are not associate with mitochondrial introns. A strong association is evident between the presence of a stand-alone endonuclease gene and high sequence diversity at the end of the endonuclease-adjacent gene (Wu and Hao, 2019). This finding is consistent with the notion that the recognition sites of endonuclease are mutation hotspots.

THE REBIRTH OF GC-CLUSTERS

The *var1* gene (or called *rps3* in non-yeast fungi) is another well-studied example of unidirectional inheritance in *S. cerevisiae* (Strausberg and Butow, 1981). The *var1* gene is polymorphic, and different forms of the *var1* gene differ by in-frame insertions of short GC-rich palindromic cluster (GC-cluster) in the coding region (Hudspeth et al., 1984). The terminal sequences of most GC clusters are repeats and different GC clusters can share the same terminal repeats. For instance, AG dinucleotide and TAG trinucleotide repeats are common among GC-clusters in *S. cerevisiae* (Weiller et al., 1989). The terminal repeats are regarded as target-site duplication, and GC-clusters have been suggested to bear ribozyme activity, which catalyzes self-cleavage and ligation reactions (Weiller et al., 1989; Lang et al., 2014). GC-cluster sequences rapidly accumulate substitutions especially in the loop regions, and also undergo dynamic merger and shuffling to form new GC-clusters (Wu and Hao, 2015). Changes at nucleotide bases as well as sequence structure result in highly variable GC cluster sequences among different yeast mitogenomes. GC-clusters are most often found in intergenic regions, but many of them are transcribed into RNAs (Wu and Hao, 2015). All these support the notion that GC clusters are transposable elements. GC-clusters can also be found in protein-coding regions. In yeast *Magnusiomyces capitatus* (in the Dipodascaceae family), GC-clusters inserted in protein-coding regions are transcribed in mRNAs, but the GC-cluster region in mRNA gets bypassed (or ignored) during translation (Lang et al., 2014). GC clusters have been suggested as recombination hotspots in mitogenomes (Dieckmann and Gandy, 1987). GC cluster–mediated gene conversion can insert or delete large genomic fragments (Weiller et al., 1991), which ultimately lead to alteration of genome size. GC clusters have been suggested to induce long AT-rich sequences into the *Nakaseomyces bacillisporus* mitogenome (Bouchier et al., 2009).

MITOCHONDRIAL DNA RECOMBINATION

Pioneer studies on mtDNA recombination through mating between *S. cerevisiae* strains can be traced back to the early 1970s (Kleese et al., 1972; Shannon et al., 1972). Recent large-scale genomic survey confirms frequent mtDNA recombination in natural *S. cerevisiae* populations (De Chiara et al., 2020). Surprisingly, yeasts were not among the organisms in early discoveries of mtDNA recombination between different species, as shown in plants (Hao et al., 2010; Mower et al., 2010). Subsequent analysis on closely related yeast mitogenomes found extensive recombination throughout the mitogenome between yeast species (Wu et al., 2015). There are two important identified issues in assessing the extent of mtDNA recombination events. (1) Accurate detection of mtDNA recombination relies on the abundance of closely related mitogenomes, preferably, an abundant number of intraspecific mitogenomes from several related species. (2) Many mtDNA recombinant events are fine-scale and often overlooked when using whole genes as the unit of phylogenetic analysis.

Fritsch et al. (2014) have constructed the genome-wide map of mtDNA recombination events in *S. cerevisiae* and found that recombinant hotspots are preferentially localized in intergenic and intronic regions. They further investigated the impact of individual depletion of four genes [namely *Ntg1* (Ling et al., 2007), *Mgt1* (Lockshon et al., 1995), *MHR1*, and *Din7* (Ling et al., 2013)] previously associated with mtDNA recombination. The deletion of *Ntg1*, *Mgt1*, and *MHR1* had little influence on mtDNA recombination hotspots, and the deletion of *Din7* resulted primarily in DNA degradation. These findings hint that the study of only nuclear-encoded genes is unlikely to achieve a complete understanding of the localization of recombination hotspots along the yeast mitogenome. Given the dynamic nature of mitochondrial -encoded sequences, such as introns and GC-clusters, a plausible alternative could be that mtDNA recombination hotspots are primarily driven by the sequence features of the mitogenome. Future studies are needed to test this hypothesis.

Mitochondrial DNA recombination can impact mitochondrial function in different ways. As described above (on GC-clusters), Non-allelic mtDNA recombination can lead to deletion and insertion of sequences. Allelic mtDNA recombination, on the other hand, can prevent mtDNA deletion and repair mitogenome damage (Ling et al., 2019). Mitochondrial recombination introduces rapid sequence changes, and could have a significant functional impact on the host. There are direct functional impacts of the introduced mito-genotype, effects of the altered mito-nuclear interaction, and effects of the interaction between the introduced mito-loci and native mito-loci (mito-mito interaction) (Wolters et al., 2018). Mitochondrial recombination can enhance phenotypic variation among diploid hybrids, and facilitate the phenotypic differentiation of hybrid species (Leducq et al., 2017). Given the prevalence of yeast mtDNA recombination and even across different species, much phenotypic diversity in yeast could have resulted from mitochondrial recombination.

YEAST MITOGENOMES ARE HIGHLY VARIABLE IN SIZE

In the Saccharomycetaceae family, the smallest and largest mitogenomes belong in the same genus *Nakaseomyces* (**Figure 1**). The mitogenome in *Candida glabrata* is 20.1 kb (Koszul et al., 2003), while the mitogenome in *Nakaseomyces bacillisporus* is over five-time larger, at 107.1 Kb (Bouchier et al., 2009). In contrast, the sizes of nuclear genomes among the *Nakaseomyces* species are remarkably similar, ranging from 10.2 to 12.3 Mb (Gabaldon et al., 2013). In the sister family Saccharomycodaceae, *Hanseniaspora uvarum* has a linear mitogenome at 11.1 kb (Pramateftaki et al., 2006), and *Saccharomycodes ludwigii* has a circular mitogenome at 69.0 kb (Nguyen et al., 2020b). Mitogenome sizes also vary among strains within the same yeast species. Among the 109 *S. cerevisiae* mitogenomes examined in

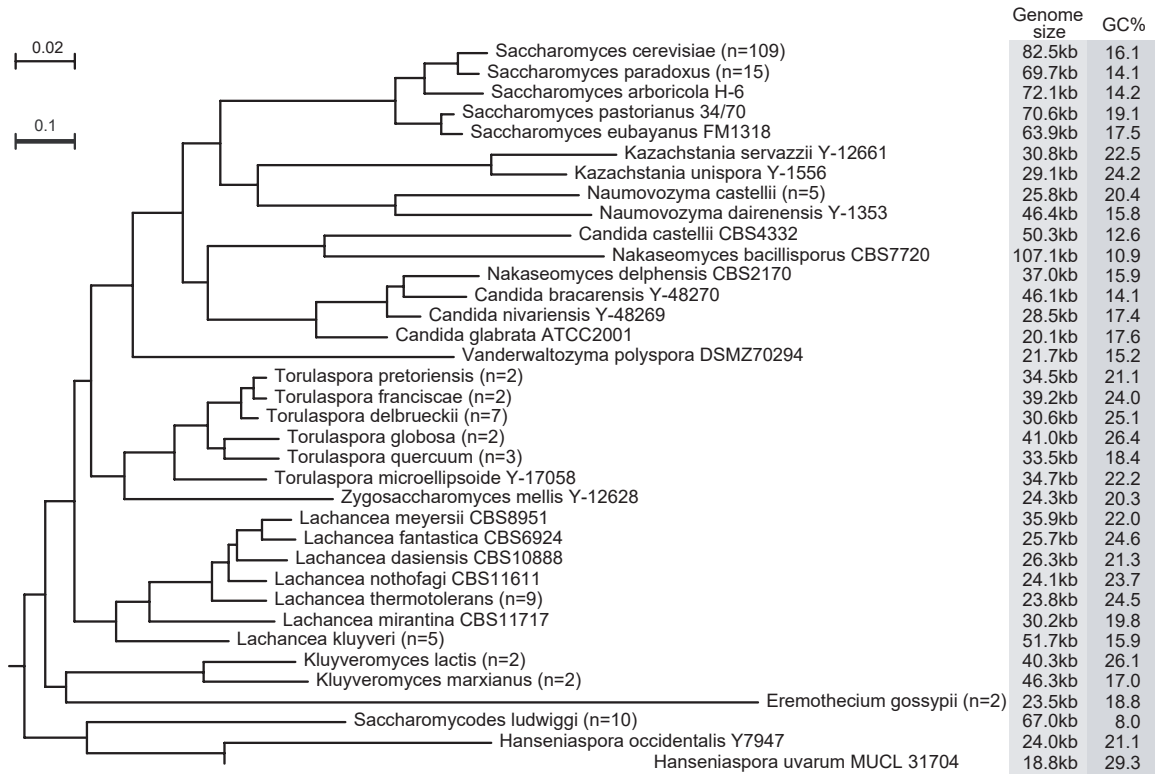


FIGURE 1 | Variation in size, sequence divergence, and GC-content of mtDNAs of 36 yeast species. The phylogenetic tree is based on nucleotide sequences of seven core mitochondrial genes (*cob*, *cox1*, *cox2*, *cox3*, *atp6*, *atp8*, and *atp9*). The very long branch leading to *Hanseniaspora uvarum* is shown by a thickened line at fivefold reduced scale. If two or more strains are analyzed in a species, the averages are shown for genome size and GC% (strain number in parentheses). Otherwise, the strain names are shown. Details of all strain names in the Saccharomycetaceae family are in Xiao et al. (2017), and the *Saccharomycodes ludwigii* strains are in Nguyen et al. (2020b).

Xiao et al. (2017), and their genome sizes range from 74.2 to 92.2 kb, The fast increasing number of yeast mitogenomes will only expand the range in mitogenome size difference.

STABLE GENE CONTENT IN YEAST MITOGENOMES

Despite the variable sizes, genes encoded in yeast mitogenomes are remarkably stable. Unlike mitogenomes in many other fungal species, none of the mitogenomes in Saccharomycetaceae and Saccharomycodaceae encode the respiratory-chain NADH dehydrogenase (complex I) (Dujon, 2010; Freel et al., 2014). The loss of complex I in yeasts is generally believed as a result of adaptation to fermentative lifestyles, where complex I is not essential (Schikora-Tamarit et al., 2021). All complete mitogenomes in Saccharomycetaceae encode eight protein genes, seven respiratory-chain protein genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, and *cox3*) and one ribosomal protein gene *var1*, SSU and LSU rRNAs, and 22–24 tRNAs. This is in contrast with a recent report that "no gene is universally conserved in fungal mitogenomes" (Fonseca et al., 2021). Although it is inevitable that fewer mitochondrial genes are shared when more diverse lineages are included (Roger et al., 2017), the variation of mitogenome sequence quality could have been an important reason for the discrepancy on gene conservation among yeast mitogenomes. For instance, the two *S. cerevisiae* mitogenomes (accessions: CM002421 and CP046458) reportedly missing *atp6* by Fonseca et al. (2021) are 22,149 bp (including 13,247 Ns for gaps) and 49,451 bp (517 Ns) in length, respectively, much shorter than the average 82.5 Kb [ranging from 74.2 to 92.2 kb in Xiao et al. (2017)] of *S. cerevisiae* mitogenomes (**Figure 1**). The Saccharomycodaceae family has a sole case of gene loss involving a protein gene. That is the *var1* gene absent from *Hanseniaspora uvarum* (Pramateftaki et al., 2006). This is likely a lineage-specific gene loss, as *var1* is present in *Saccharomycodes ludwigii* (Nguyen et al., 2020b), a related species within the Saccharomycodaceae family.

FACTORS DRIVING MITOGENOME SIZE VARIATION

Mobile introns and variable intergenic regions are known factors driving variation of yeast mitogenome architecture (Bouchier et al., 2009; Freel et al., 2014). Ironically, the mitogenome in *Nakaseomyces bacillisporus* at 107 kb contains no intron, while

mitogenomes in *Candida glabrata* at around 20 kb contain 3–4 introns in at least four intron-distribution patterns [(Koszul et al., 2003), and unpublished observation]. This irony can be solved by separate analyses at different time scales (Xiao et al., 2017). When intraspecific mitogenomes (within the same species) are compared, intron sequences show the highest variance in length and significantly overrepresented in large mitogenomes. When interspecific mitogenomes (among different species) are compared, tandem repeats show the highest variance in sequence length and significantly overrepresented in large mitogenomes. In other words, the rapid turnover of mobile introns can significantly impact genome size, but the number of available introns insertion sites are limited; while expansion and contraction of repeats may cause only subtle change per event, but they take place persistently with little space limit.

ORIGIN OF MITOGENOME SIZE

The question on whether genome size is under selective constraint has been a subject of debate (Lynch and Conery, 2003; Whitney and Garland, 2010; Shtolz and Mishmar, 2019). The mutational burden hypothesis (MBH) was postulated to explain the origin of organellar genome size (Lynch et al., 2006). Introns and intergenic DNAs in mitogenomes are genetic liability, as they are targets for deleterious and potentially lethal mutations (Lynch et al., 2006). Following MBH, introns and intergenic DNAs tend to accumulate when natural selection is less efficient at purging hazardous non-coding DNA. Using the ratio of non-synonymous over synonymous Ka/Ks rates as a proxy for the level of genetic drift, a significant positive correlation was evident between the genome-wide Ka/Ks ratios and mitogenome sizes among seven yeast species with sufficient intraspecific diversity (Xiao et al., 2017). This finding is consistent with the notion that introns, GC-clusters, and repeats in yeast mitogenomes are mostly deleterious (Bernardi, 2005; Wu and Hao, 2014, 2015). *Hanseniaspora uvarum* has accelerated sequence evolution compared with related species (**Figure 1**) and has a small genome size, which is consistent with the MBH hypothesis. Genetic drift still faces challenges to explain mitogenome size variation in many yeast species. For example, significant relaxation of mitochondrial functions has been documented after whole-genome duplication (Jiang et al., 2008), yet many post-whole-genome duplication species (including *Candida glabrata*) have compact mitogenomes. Unfortunately, many yeast species either lack intraspecific genomic data or suffer insufficient sequence diversity, making it impossible to estimate the degree of genetic drift. To address this issue, extended efforts are needed to sample and sequence an abundant number of intraspecific strains in a broad range of yeast species.

EXTREME GENOME-WIDE G + C CONTENT

Even though yeast mitogenomes are overall AT-rich, their G + C-contents vary greatly. GC-content of the *Hanseniaspora uvarum* mitogenome is 29.3% (Pramateftaki et al., 2006), while the most AT-rich mitogenome in *Saccharomycodes ludwigii* is at 7.6% G + C (Nguyen et al., 2020b). Generally speaking, mutation is nearly universally biased from C/G to T/A (Hershberg and Petrov, 2010), and mutation rates are often higher at C/G nucleotides than at A/T nucleotides (Zhu et al., 2014). Alterations of GC content have been shown to impact mutation and recombination rates (Kiktev et al., 2018). The mitogenomes with extreme base composition offer important and unique insights into the mechanisms governing mutation processes (Gardner et al., 2002; McCutcheon and Moran, 2010; Smith et al., 2011; Su et al., 2019). Comparative genomics of the 10 extreme AT-rich mitogenomes in *Saccharomycodes ludwigii* (Nguyen et al., 2020b) found a strong mutation bias toward A/T, but the expected equilibrium G + C content under mutation pressure alone is still higher than observed G + C content. Interestingly, mitogenomes in *Saccharomycodes ludwigii* undergo frequent recombination, a genetic process that normally increases G + C content by GC-biased gene conversion (Pessia et al., 2012). These findings suggest other mechanisms alongside with AT-biased mutation operating to increase A/T in *Saccharomycodes ludwigii*. Another important, but perhaps underappreciated finding is the prevalence of indel mutations in yeast mitogenomes (Xiao et al., 2017). Indel mutations contribute much more to genomic variation among closely related mitogenomes than nucleotide substitutions. Further studies are needed to investigate the molecular mechanisms driving indel mutations and quantitatively model the evolutionary process of indel mutations.

VARIABLE EVOLUTIONARY RATES AMONG YEAST MITOGENOMES

Yeast mitogenomes show variable evolutionary rates. The branch leading to *Hanseniaspora uvarum* is at least five times longer than the branch leading to its sister species. Similarly, the branch leading to a plant pathogen *Eremothecium gossypii* is at least three times longer than the branch leading to the related *Kluyveromyces* species (**Figure 1**). Spontaneous mitochondrial mutation rates have been measured in several yeast species. The mitochondrial base-substitution mutation (BSM) rates are all higher than their corresponding nuclear BSM rates. The mitochondrial BSM rates in *S. cerevisiae* range from 4.47×10^{-10} (Sharp et al., 2018) to 122.3×10^{-10} per site per cell division (Lynch et al., 2008). Please also note that the *S. cerevisiae* FY10 strain (isogenic to S288c) used in the Lynch et al. (2008) study contains a single non-synonymous mutation in the mtDNA polymerase (*MIP1*) linked to reduced fidelity of mtDNA replication (Baruffini et al., 2007b). *Hanseniaspora uvarum* does have a higher mitochondrial BSM rate than other two *Hanseniaspora* species, at 13.1×10^{-10} mutations per site per cell division, compared with 5.94×10^{-10} mutations per site per cell division in *Hanseniaspora valbyensis*, and 3.65×10^{-10} mutations per site per cell division in *Hanseniaspora osmophila* (Nguyen et al., 2020a). The measured spontaneous mutation rates will also help us estimate the effective population size for each species following equation

$\pi_{silent} = 2N_e \mu$, where π_{silent} is nucleotide diversity at silent sites, N_e is effective population size, μ is mutation rate (Lynch, 2006). Precise estimation of effective population size for mtDNA holds the key to understanding the significance of mtDNA recombination at the population level. To achieve this, it is critical to obtain both intraspecific diversity and spontaneous mutation rate for a variety of yeast species. Extra attention must also be paid to accumulate sufficient number of mitochondrial mutations due to the small mitogenome size relative to the nuclear genome size.

MOVING FORWARD

Yeast mitogenomes are highly diverse ranging from fast-evolving compact mitogenomes (similar to animal mitogenomes) to slow-evolving mitogenomes inflated by large introns, repeats and non-coding sequences (similar to plant mitogenomes). The fast-growing yeast mitogenome data have allowed us to begin to identify mechanisms driving genome diversity. Specific efforts are needed to sequence and study an abundant number of intraspecific strains from closely related species for contrast genomic features. The future of yeast mitogenome studies is bright, and the generated knowledge will no doubt benefit our understanding of mitogenomes much beyond the fungal kingdom.

AUTHOR CONTRIBUTIONS

WH wrote the manuscript.

FUNDING

Previous work in the lab was supported by funds from Wayne State University and the National Science Foundation (ISO1838291) to WH.

ACKNOWLEDGMENTS

I am grateful to GH and three reviewers for helpful comments on a previous version of this manuscript.

REFERENCES

Baruffini, E., Ferrero, I., and Foury, F. (2007a). Mitochondrial DNA defects in *Saccharomyces cerevisiae* caused by functional interactions between DNA polymerase gamma mutations associated with disease in human. *Biochim. Biophys. Acta* 1772, 1225–1235. doi: 10.1016/j.bbadis.2007.10.002

Baruffini, E., Lodi, T., Dallabona, C., and Foury, F. (2007b). A single nucleotide polymorphism in the DNA polymerase gamma gene of *Saccharomyces cerevisiae* laboratory strains is responsible for increased mitochondrial DNA mutability. *Genetics* 177, 1227–1231. doi: 10.1534/genetics.107.079293

Bazin, E., Glemin, S., and Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572. doi: 10.1126/science.1122033

Bernardi, G. (2005). Lessons from a small, dispensable genome: the mitochondrial genome of yeast. *Gene* 354, 189–200. doi: 10.1016/j.gene.2005.03.024

Borst, P., and Grivell, L. A. (1978). The mitochondrial genome of yeast. *Cell* 15, 705–723. doi: 10.1016/0092-8674(78)90257-x

Bouchier, C., Ma, L., Creno, S., Dujon, B., and 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, 1283–1292. doi: 10.1111/j.1567-1364.2009.00551.x

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716.

Butow, R. A. (1985). Nonreciprocal exchanges in the yeast mitochondrial genome. *Trends Genet.* 1, 81–84.

Cho, Y., and Palmer, J. D. (1999). Multiple acquisitions *via* horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol. Biol. Evol.* 16, 1155–1165.

Christensen, A. C. (2013). Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol. Evol.* 5, 1079–1086. doi: 10.1093/gbe/evt069

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. doi: 10.1186/s12915-020-00786-4

Dieckmann, C. L., and Gandy, B. (1987). Preferential recombination between GC clusters in yeast mitochondrial DNA. *EMBO J.* 6, 4197–4203.

Dimitrov, L. N., Brem, R. B., Kruglyak, L., and Gottschling, D. E. (2009). Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics* 183, 365–383. doi: 10.1534/genetics.109.104497

Dujon, B. (1989). Group I introns as mobile genetic elements: facts and mechanistic speculations–a review. *Gene* 82, 91–114.

Dujon, B. (2010). Yeast evolutionary genomics. *Nat. Rev. Genet.* 11, 512–524. doi: 10.1038/nrg2811

Dujon, B. (2020). Mitochondrial genetics revisited. *Yeast* 37, 191–205. doi: 10.1002/yea.3445

Dujon, B., Slonimski, P. P., and Weill, L. (1974). Mitochondrial genetics IX: a model for recombination and segregation of mitochondrial genomes in saccharomyces cerevisiae. *Genetics* 78, 415–437. doi: 10.1093/genetics/78.1.415

Fekete, V., Cierna, M., Poláková, S., Piskur, J., and Sulo, P. (2007). Transition of the ability to generate petites in the *Saccharomyces*/*Kluyveromyces* complex. *FEMS Yeast Res.* 7, 1237–1247. doi: 10.1111/j.1567-1364.2007.00287.x

Fonseca, P. L. C., De-Paula, R. B., Araújo, D. S., Tomé, L. M. R., Mendes-Pereira, T., Rodrigues, W. F. C., et al. (2021). Global characterization of fungal mitogenomes: new insights on genomic diversity and dynamism of coding genes and accessory elements. *Front. Microbiol.* 12:787283. doi: 10.3389/fmicb.2021.787283

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

Fritsch, E. S., Chabbert, C. D., Klaus, B., and Steinmetz, L. M. (2014). A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 198, 755–771. doi: 10.1534/genetics.114.166637

Gabaldon, T., Martin, T., Marcet-Houben, M., Durrens, P., Bolotin-Fukuhara, M., Lespinet, O., et al. (2013). Comparative genomics of emerging pathogens in the *Candida glabrata* clade. *BMC Genomics* 14:623. doi: 10.1186/1471-2164-14-623

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511. doi: 10.1038/nature01097

Goddard, M. R., and Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13880–13885.

Hao, W., Richardson, A. O., Zheng, Y., and Palmer, J. D. (2010). Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21576–21581. doi: 10.1073/pnas.1016295107

Hershberg, R., and Petrov, D. A. (2010). Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.* 6:e1001115. doi: 10.1371/journal.pgen.1001115

Hudspeth, M. E., Vincent, R. D., Perlman, P. S., Shumard, D. S., Treisman, L. O., and Grossman, L. I. (1984). Expandable var1 gene of yeast mitochondrial DNA:

in-frame insertions can explain the strain-specific protein size polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* 81, 3148–3152. doi: 10.1073/pnas.81.10.3148

Jacquier, A., and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 41, 383–394. doi: 10.1016/s0092-8674(85)80011-8

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

Kiktev, D. A., Sheng, Z., Lobachev, K. S., and Petes, T. D. (2018). GC content elevates mutation and recombination rates in the yeast Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* 115, E7109–E7118. doi: 10.1073/pnas.1807334115

Kleese, R. A., Grotbeck, R. C., and Snyder, J. R. (1972). Recombination among three mitochondrial genes in yeast (*Saccharomyces cerevisiae*). *J. Bacteriol.* 112, 1023–1025.

Koszul, R., Malpertuy, A., Frangeul, L., Bouchier, C., Wincker, P., Thierry, A., et al. (2003). The complete mitochondrial genome sequence of the pathogenic yeast *Candida* (*Torulopsis*) *glabrata*. *FEBS Lett.* 534, 39–48.

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397. doi: 10.1146/annurev.genet.33.1.351

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

Leducq, J. B., Henault, M., Charron, G., Nielly-Thibault, L., Terrat, Y., Fiumera, H. L., et al. (2017). Mitochondrial recombination and introgression during speciation by hybridization. *Mol. Biol. Evol.* 34, 1947–1959. doi: 10.1093/molbev/msx139

Ling, F., Bradshaw, E., and Yoshida, M. (2019). Prevention of mitochondrial genomic instability in yeast by the mitochondrial recombinase Mhr1. *Sci. Rep.* 9:5433. doi: 10.1038/s41598-019-41699-9

Ling, F., Hori, A., and Shibata, T. (2007). DNA recombination-initiation plays a role in the extremely biased inheritance of yeast [rho-] mitochondrial DNA that contains the replication origin ori5. *Mol. Cell Biol.* 27, 1133–1145. doi: 10.1128/MCB.00770-06

Ling, F., Hori, A., Yoshitani, A., Niu, R., Yoshida, M., and Shibata, T. (2013). Din7 and Mhr1 expression levels regulate double-strand-break-induced replication and recombination of mtDNA at ori5 in yeast. *Nucleic Acids Res.* 41, 5799–5816. doi: 10.1093/nar/gkt273

Lockshon, D., Weifel, S. G., Freeman-Cook, L. L., Lorimer, H. E., Brewer, B. J., and Fangman, W. L. (1995). A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* 81, 947–955.

Lynch, M. (2006). The origins of eukaryotic gene structure. *Mol. Biol. Evol.* 23, 450–468. doi: 10.1093/molbev/msj050

Lynch, M., and Conery, J. S. (2003). The origins of genome complexity. *Science* 302, 1401–1404. doi: 10.1126/science.1089370

Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. *Science* 311, 1727–1730. doi: 10.1126/science.1118884

Lynch, M., Ung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., et al. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9272–9277. doi: 10.1073/pnas.0803466105

McCutcheon, J. P., and Moran, N. A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol. Evol.* 2, 708–718. doi: 10.1093/gbe/evq055

Mower, J. P., Stefanović, S., Hao, W., Gummow, J. S., Jain, K., Ahmed, D., et al. (2010). Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.* 8:150. doi: 10.1186/1741-7007-8-150

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

Nguyen, D. T., Wu, B., Long, H., Zhang, N., Patterson, C., Simpson, S., et al. (2020a). Variable spontaneous mutation and loss of heterozygosity among heterozygous genomes in yeast. *Mol. Biol. Evol.* 37, 3118–3130. doi: 10.1093/molbev/msaa150

Nguyen, D. T., Wu, B., Xiao, S., and Hao, W. (2020b). 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

Perlman, P. S., and Butow, R. A. (1989). Mobile introns and intron-encoded proteins. *Science* 246, 1106–1109. doi: 10.1126/science.2479980

Pessia, E., Popa, A., Mousset, S., Rezvoy, C., Duret, L., and Marais, G. A. (2012). Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biol. Evol.* 4, 675–682. doi: 10.1093/gbe/evs052

Piskur, J. (2001). Origin of the duplicated regions in the yeast genomes. *Trends Genet.* 17, 302–303. doi: 10.1016/s0168-9525(01)02308-3

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

Sanchez-Puerta, M. V., Abbona, C. C., Zhuo, S., Tepe, E. J., Bohs, L., Olmstead, R. G., et al. (2011). Multiple recent horizontal transfers of the cox1 intron in Solanaceae and extended co-conversion of flanking exons. *BMC Evol. Biol.* 11:277. doi: 10.1186/1471-2148-11-277

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

Shannon, C., Rao, A., Douglass, S., and Criddle, R. S. (1972). Recombination in yeast mitochondrial DNA. *J. Supramol. Struct.* 1, 145–152. doi: 10.1002/jss.400010207

Shao, R., Kirkness, E. F., and Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. *Genome Res.* 19, 904–912. doi: 10.1101/gr.083188.108

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

Shtolz, N., and Mishmar, D. (2019). The mitochondrial genome–on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Front. Ecol. Evol.* 7:342. doi: 10.3389/fevo.2019.00342

Sloan, D. B., Alverson, A. J., Chuckalovcak, J. P., Wu, M., McCauley, D. E., Palmer, J. D., et al. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 10:e1001241. doi: 10.1371/journal.pbio.1001241

Smith, D. R. (2015). The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief. Funct. Genomics* 15, 47–54. doi: 10.1093/bfgp/elv027

Smith, D. R., Burki, F., Yamada, T., Grimwood, J., Grigoriev, I. V., Van Etten, J. L., et al. (2011). The GC-rich mitochondrial and plastid genomes of the green alga *Coccomyxa* give insight into the evolution of organelle DNA nucleotide landscape. *PLoS One* 6:e23624. doi: 10.1371/journal.pone.0023624

Smith, D. R., Kayal, E., Yanagihara, A. A., Collins, A. G., Pirro, S., and Keeling, P. J. (2012). First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. *Genome Biol. Evol.* 4, 52–58. doi: 10.1093/gbe/evr127

Strausberg, R. L., and Butow, R. A. (1981). Gene conversion at the var1 locus on yeast mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 78, 494–498.

Strope, P. K., Skelly, D. A., 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

Su, H. J., Barkman, T. J., Hao, W., Jones, S. S., Naumann, J., Skippington, E., et al. (2019). Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 934–943. doi: 10.1073/pnas.1816822116

Weiller, G. F., Bruckner, H., Kim, S. H., Pratje, E., and Schweyen, R. J. (1991). A GC cluster repeat is a hotspot for mit- macro-deletions in yeast mitochondrial DNA. *Mol Gen Genet* 226, 233–240.

Weiller, G., Schueller, C. M., and Schweyen, R. J. (1989). Putative target sites for mobile G + C rich clusters in yeast mitochondrial DNA: single elements and tandem arrays. *Mol. Gen. Genet.* 218, 272–283.

Whitney, K. D., and Garland, T. Jr. (2010). Did genetic drift drive increases in genome complexity? *PLoS Genet.* 6:e1001080. doi: 10.1371/journal.pgen.1001080

Whittaker, P. A., Hammond, R. C., and Luha, A. A. (1972). Mechanism of mitochondrial mutation in yeast. *Nat. New Biol.* 238, 266–268. doi: 10.1038/newbio238266a0

Wilkie, D., and Thomas, D. Y. (1973). Mitochondrial genetic analysis by zygote cell lineages in *Saccharomyces cerevisiae*. *Genetics* 73, 367–377. doi: 10.1093/genetics/73.3.367

Wolters, J. F., Charron, G., Gaspary, A., Landry, C. R., Fiumera, A. C., and Fiumera, H. L. (2018). Mitochondrial recombination reveals mito-mito epistasis in yeast. *Genetics* 209, 307–319. doi: 10.1534/genetics.117.300660

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

Wu, B., and Hao, W. (2015). A dynamic mobile DNA family in the yeast mitochondrial genome. *G3* 5, 1273–1282. doi: 10.1534/g3.115.017822

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

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

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

Yin, S., Heckman, J., and RajBhandary, U. L. (1981). Highly conserved GC-rich palindromic DNA sequences flank tRNA genes in Neurospora crassa mitochondria. *Cell* 26, 325–332. doi: 10.1016/0092-8674(81)90201-4

Zhu, Y. O., Siegal, M. L., Hall, D. W., and Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2310–E2318. doi: 10.1073/pnas.1323011111

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Fungal Biology

ORIGINAL RESEARCH
published: 09 February 2022
doi: 10.3389/ffunb.2021.802511

The First Mitochondrial Genome of *Ciborinia camelliae* and Its Position in the Sclerotiniaceae Family

Irene Valenti, Luca Degradi, Andrea Kunova, Paolo Cortesi, Matias Pasquali and Marco Saracchi*

Department of Food, Environmental and Nutritional Sciences, University of Milan, Milan, Italy

Ciborinia camelliae is the causal agent of camellia flower blight (CFB). It is a hemibiotrophic pathogen, inoperculate Discomycete of the family Sclerotiniaceae. It shows host and organ specificity infecting only flowers of species belonging to the genus *Camellia*, causing serious damage to the ornamental component of the plant. In this work, the first mitochondrial genome of *Ciborinia camellia* is reported. The mitogenome was obtained by combining Illumina short read and Nanopore long read technology. To resolve repetitive elements, specific primers were designed and used for Sanger sequencing. The manually curated mitochondrial DNA (mtDNA) of the Italian strain DSM 112729 is a circular sequence of 114,660 bp, with 29.6% of GC content. It contains two ribosomal RNA genes, 33 transfer RNAs, one RNase P gene, and 62 protein-coding genes. The latter include one gene coding for a ribosomal protein (*rps3*) and the 14 typical proteins involved in the oxidative metabolism. Moreover, a partial mtDNA assembled from a contig list was obtained from the deposited genome assembly of a New Zealand strain of *C. camelliae*. The present study contributes to understanding the mitogenome arrangement and the evolution of this phytopathogenic fungus in comparison to other Sclerotiniaceae species and confirms the usefulness of mitochondrial analysis to define phylogenetic positioning of this newly sequenced species.

Keywords: fungal diversity, pathogen of *Camellia*, mitochondrial assembly and annotation, short and long read sequencing, mitochondrial diversity

OPEN ACCESS

Edited by:
Georg Hausner,
University of Manitoba, Canada

Reviewed by:
Daniel S. Araújo,
Loyola University Chicago, United States
Hsin-Han Lee,
Academia Sinica, Taiwan

****Correspondence:***
Matias Pasquali
matias.pasquali@unimi.it

Specialty section:
This article was submitted to Fungal Genomics and Evolution, a section of the journal Frontiers in Fungal Biology

Received: *26 October 2021*
Accepted: *06 December 2021*
Published: *09 February 2022*

Citation:
Valenti I, Degradi L, Kunova A, Cortesi P, Pasquali M and Saracchi M (2022) The First Mitochondrial Genome of Ciborinia camelliae and Its Position in the Sclerotiniaceae Family. Front. Fungal Biol. 2:802511. doi: 10.3389/ffunb.2021.802511

INTRODUCTION

The genus *Ciborinia* involves 23 different species of host-specific pathogens (https://eol.org/). *Ciborinia* spp. are members of the Sclerotiniaceae family (order Helotiales; phylum Ascomycota), which includes 14 genera of phytopathogenic fungi with a wide host range. The Sclerotiniaceae's ability to infect different hosts and adapt to various environments results in serious economic damage (Bolton et al., 2006). Unlike the other Sclerotiniaceae, *Ciborinia camelliae* Kohn infects only plants of the genus *Camellia* causing the camellia flower blight (CFB) (Taylor and Long, 2000; Saracchi et al., 2019). Sclerotia in fallen flowers lie dormant on the ground or in plant debris over summer, autumn, and winter. Toward the end of winter, the sclerotia begin to germinate, producing apothecia from which ascospores are released. The pathogen infects and colonizes only flowers, where brown spots appear and spread to the entire organ. Flowers tend to fall prematurely. This causes important damage to the camellia floriculture industry (Taylor and Long, 2000; Denton-Giles et al., 2013).

Studies on the diversity within the genus *Ciborinia* are scarce (Kohn, 1979). The *C. camelliae* variability was investigated using the UP-PCR analysis on strains from America and New Zealand, demonstrating a relatively low level of genetic variations within the two populations (van Toor et al., 2005). On the other hand, a recent study based on morpho-cultural characterization and ITS analysis of numerous Italian strains showed significant variability among the isolates (Saracchi et al., 2022).

The mitochondria are membrane-enclosed compartments that play a central role in providing energy to eukaryotic cells by oxidative phosphorylation (Newmeyer and Ferguson-Miller, 2003; Kriváková et al., 2005; Lv et al., 2017). They are also involved in the signal amplification leading to apoptosis (Goodsell, 2010), antifungal drug resistance as well as in virulence and pathogenicity (Shingu-Vazquez and Traven, 2011; Sandor et al., 2018; Kulik et al., 2020). The study of mitochondrial (mt) genomes reveals interesting features about the pathogen evolution and its relationships with the other related species (Ballard and Whitlock, 2004; Aguileta et al., 2014; Mardanov et al., 2014; Kulik et al., 2020). Even across distantly connected species, the mt genes are largely conserved due to their essential role in cell vitality (Medina et al., 2020). Nevertheless, intron numbers and secondary structures are highly variable and are prone to evolving rapidly. All these features make mitogenome a functional source for phylogenetic studies (Hamari et al., 2002; Burger et al., 2003; Galtier et al., 2009). The mitochondrial genome is often a circular double-stranded molecule with a condensed gene arrangement (Burger et al., 2003). Complete fungal mitogenomes vary notably in size. Typically, the fungal mitochondrial genome contains 14 conserved protein-coding genes involved in the oxidative metabolism (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5,* and *nad6*) and one ribosomal protein S3 gene (*rps3*). Additionally, each mtDNA has two ribosomal RNA (rRNA) genes for the large and small rRNA subunit (*rnl* and *rns*) and a variable number of tRNAs (Bullerwell and Lang, 2005).

In this study, we present a comprehensive analysis of the first *C. camelliae* mitochondrial DNA. The complete sequence of an Italian strain was obtained with a hybrid strategy combining Illumina Hiseq paired reads and MinIon Nanopore long reads sequencing. The mitochondrial sequences were identified and assembled together constructing the complete and circular mitogenome. The results were compared with data from the available Sclerotiniaceae mitogenomes. This study allows to position the *C. camelliae* pathogen in the Sclerotiniaceae family and offers some insight into the specificities of mtDNA of *C. camelliae*.

MATERIALS AND METHODS

Sampling, DNA Extraction, and Sequencing

The investigated strain was isolated from pieces of sclerotium collected in Oggebbio (Verbania, Italy) (45°59′47.5782″ N, 08°39′05.9659″ E) and showing *C. camelliae* characteristics. The pathogen was grown on Potato Dextrose Agar medium (PDA: 800 mL/L of potato extract; 20 g/L glucose, BioFROXX, Germany; 15 g/L agar, Applichem, Germany). The isolate was identified according to colony morphology and microscopic analysis of conidia and ascospore traits, as well as ITS sequence (Saracchi et al., 2022). The strain ITAC2 is maintained in the laboratory of Plant Pathology at the Department of Food, Environmental and Nutritional Sciences (University of Milan, Italy), and it is also deposited in a public collection (DSMZ-German Collection of Microorganisms and Cell Cultures GmbH) with the accession number DSM 112729.

High-molecular-weight genomic DNA was extracted from conidia (10^9 conidia/mL) using DNeasy PowerSoil Pro Kit (Qiagen, Hilden, Germany). DNA was quantified using the Qbit Fluorometer (Invitrogen, Thermo Fisher Scientific—USA) with the Qbit® dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific—USA). The genome quality was checked by spectrophotometer and 1% agarose gel electrophoresis. The gel was visualized using UV-transilluminator Gel Doc 2000 (BIO-RAD laboratories, USA) and Quantity One software (BIO-RAD laboratories, USA) to verify possible smearing of the DNA band.

Two different platforms were used to sequence the genomic DNA: Oxford Nanopore Technologies (ONT) and Illumina Hiseq, executed by Eurofins Genomics (Ebersberg, Germany). The ONT sequencing was performed by the MinION system (FLO-MIN-106 R9.4 flow-cell) using EXP-NBD104, EXPNBD114 in conjunction with the SQK-LSK109 kit and also SQK-RAD004 sequencing kit. A complete mitochondrial genome was obtained by combining long Nanopore reads and short Illumina reads. Sequencing data were analyzed using the European Galaxy web platform (https://usegalaxy.eu/) tools (Afgan et al., 2018) and Geneious Prime software, version 2021.1.1 (Biomatters, Auckland, New Zealand).

The whole genome assembly was performed using Flye v.2.8.3+galaxy0 (Lin et al., 2016) with a default setting. Medaka tool v.1.3.2 +galaxy0 was employed for autopolishing, and mapping of short reads on the draft assembly was performed by minimap2 on the same platform (Li, 2018). The final assembly was obtained using Pilon v.1.20.1 (Walker et al., 2014).

Mitogenome Assembly

Contigs including the mitochondrial DNA were identified within the whole genome assembly using NCBI data as reference, and BLAST or minimap2 (Li, 2018) as a tool. The detected sequences were extracted and assembled using Geneious Prime software. Single uncalled bases (Ns) and small gaps, resulting from the genome assembly, were fixed by mapping raw reads on the draft assembly and performing manual corrections. Ns and gaps were replaced with the matching bases from illumina reads. Re-mapping was performed to validate the corrections. The mitogenome was circularized by searching the terminal sequences using the minimap2 tool.

PCR Analysis

After the assembly, the mitochondrial DNA exhibited some unsure repeated regions and sets of uncalled bases (Ns). These mismatches were solved by PCR analysis, using newly designed primers. Primer3web (Untergasser et al., 2012) program was used to design and validate our primers. The new primer pairs were synthesized by Eurofins Genomics (Srl Vimodrone, Italy)

TABLE 1 | Primer sequences table.

Primer name		Primer Sequence 5′ → 3′
Primer1	Forward	TCGCGATCCATTACCATCTCT
Primer2	Reverse	CCTAAATTTCACGTGGCATGC
Primer3	Forward	AGTTGGTTGCAGTCGTTTGG
Primer4	Reverse	CAGTTTGGCACCTCGATGTC
Primer5	Reverse	TGACGGGTTTTAATCAGGGGT
Primer6	Forward	AAATGTTCCCTCTGCGTCAG
Primer7	Reverse	TCTACTAAGCGAATAGGTCCACA

(**Table 1**). The primer pairs 1–2 and 6–7 were used to validate the same sequence amplifying regions with different lengths. The Primer3 was paired with both Primer4 and Primer5. Two PCR rounds were conducted. The first round, using primers 1–2, was performed with 32 cycles in a total 25 μL reaction volume containing: 12.5 μL of Q5® Hot Start High-Fidelity 2X Master Mix (Biolabs, New England), 0.5 μM of forward and reverse primer, 1 μL of DNA sample and the remaining volume of sterile distilled water. The PCR program involved the initial denaturation at 98°C for 2′, 32 cycles of denaturation at 98°C for 20″, annealing at 66°C for 20″, extension at 72°C for 2.5′, followed by the final extension at 72°C for 7′. The second PCR round, using the primer pairs 3–4, 3–5 and 6–7, was performed in a total volume of 30 μL which contained: 0.18 μL of GoTaq® DNA Polymerase 5 U/μL (Promega, Madison, WI, USA), 6 μL of GoTaq® Reaction Buffer 5X (Promega, Madison, WI, USA), 1.2 μL of 10 mM dNTP (Promega, Madison, WI, USA), 1.5 μL of 50 μM forward primer, 1.5 μL of 50 μM reverse primer, 18.62 μL of sterile distilled water and 1 μL of DNA sample. The PCR reaction was carried out with the following parameters: 94°C for 2′; 32 cycles at 95°C for 30″, 55°C for 20″ and 72°C for 45″; 5′ at 72°C. Thermal Cyclers (iCycler-BIO-RAD laboratories, USA and VeritiPro^{TM}96-Well Thermal Cycler, Applied Biosystems by Thermo Fisher Scientific) were used to amplify the DNA regions. The reaction products were visualized by electrophoresis on a 1% agarose gel containing ethidium bromide. All PCR products were sequenced using the Sanger technology with the same primers (Eurofins Genomics, Ebersberg, Germany). The sequencing data were analyzed by Geneious Prime software. These sequences were compared to the draft of mitochondrial DNA.

Mitogenome Annotation and Characterization

The mitochondrial genes were annotated using MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) (Beck and Lang, 2010) and RNAweasel (https://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl) (Lang et al., 2007) tools and MITOS WebServer (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt et al., 2013). Manual corrections were required. To verify the annotation results, sequences and encoded proteins were compared with related species using BLASTp searches against the NCBI database. For each open reading frame (ORF), the protein with the highest similarity was found among biological sequences in the NCBI database. These results were further verified by BLASTn analysis.

Comparative Analysis

The mitogenome features were compared among some closely related species in the Sclerotiniaceae family, order Helotiales (*Botryotinia fuckeliana* (KC 832409), *Ciboria shiraiana* (CM 017871.1), *C. camelliae* (GCA_001247705.1), *Monilinia fructicola* (NC_056195.1), *Monilinia laxa* (NC_051483.1), *Monilinia polystroma* (GCA_002909645.1), *Sclerotinia borealis* (KJ434027) and *Sclerotinia sclerotiorum* (KT283062). *Glarea lozoyensis* (Order Helotiales; Family Helotiaceae) was used as an outgroup. For all investigated species, data were downloaded from the NCBI database. In species with no mitochondrial genome available, we used BLAST searches to identify the mtDNA in the total genome assembly and subsequently, each detected mitochondrial sequence was annotated according to the procedure described earlier. When possible, these additional results were used in comparative analyses. Based on data availability, seven Sclerotiniaceae species (*Botryotinia fuckeliana, Ciboria shiraiana, C. camelliae* strain ITAC2*, M. fructicola, M. laxa, S. borealis,* and *S. sclerotiorum*) were investigated in terms of gene arrangement. *Monilinia polystroma* and *C. camelliae* ICMP 19812 strain were excluded from this analysis due to the lack of the whole mitochondrial sequence. Multiple mitogenome alignment was performed within the previously cited Sclerotiniaceae using the MAUVE tool of Geneious Prime software and setting *nad4L* gene as starting point. MAUVE alignment identifies homologous regions shared by two or more mitogenomes. These regions are denominated locally collinear blocks (LCBs) (Darling et al., 2004). Tandem Repeat Finder was employed to identify tandem repeats within the mitochondrial genomes (Benson, 1999). The codon usage was estimated using both Sequence Manipulation Suite webserver (Stothard, 2000) and cusp tool of the European Galaxy web platform (https://galaxy-iuc.github.io/emboss-5.0-docs/cusp.html).

The phylogenetic study was performed on the multiple alignments of 14 concatenated mitochondrial proteins using amino acid sequences obtained from each mtDNA. A maximum-likelihood phylogenetic tree was executed by IQ-TREE (Trifinopoulos et al., 2016) web server selecting default setting, and performed by MEGAX software (Kumar et al., 2018).

RESULTS

The circular mitochondrial genome of *C. camelliae* strain ITAC2, submitted to the GenBank database with the accession number OK326902, has a total length of 114,660 bp. The nucleotide composition is the following: 34.7% of A, 13.2% of C, 16.5% of G, 35.7% of T with a GC content of 29.6%.

This mitogenome contains all 14 typical genes encoding the subunits of ATP synthase (*atp6*, *atp8*, and *atp9*), NADH dehydrogenase (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*), apocytochrome b (*cob*), and cytochrome c oxidase (*cox1*, *cox2,* and *cox3)*. Additionally, the conserved ribosomal protein-coding gene S3 (*rps3*), untranslated genes of the large and small ribosomal RNAs (*rnl* and *rns*), the ribonuclease P

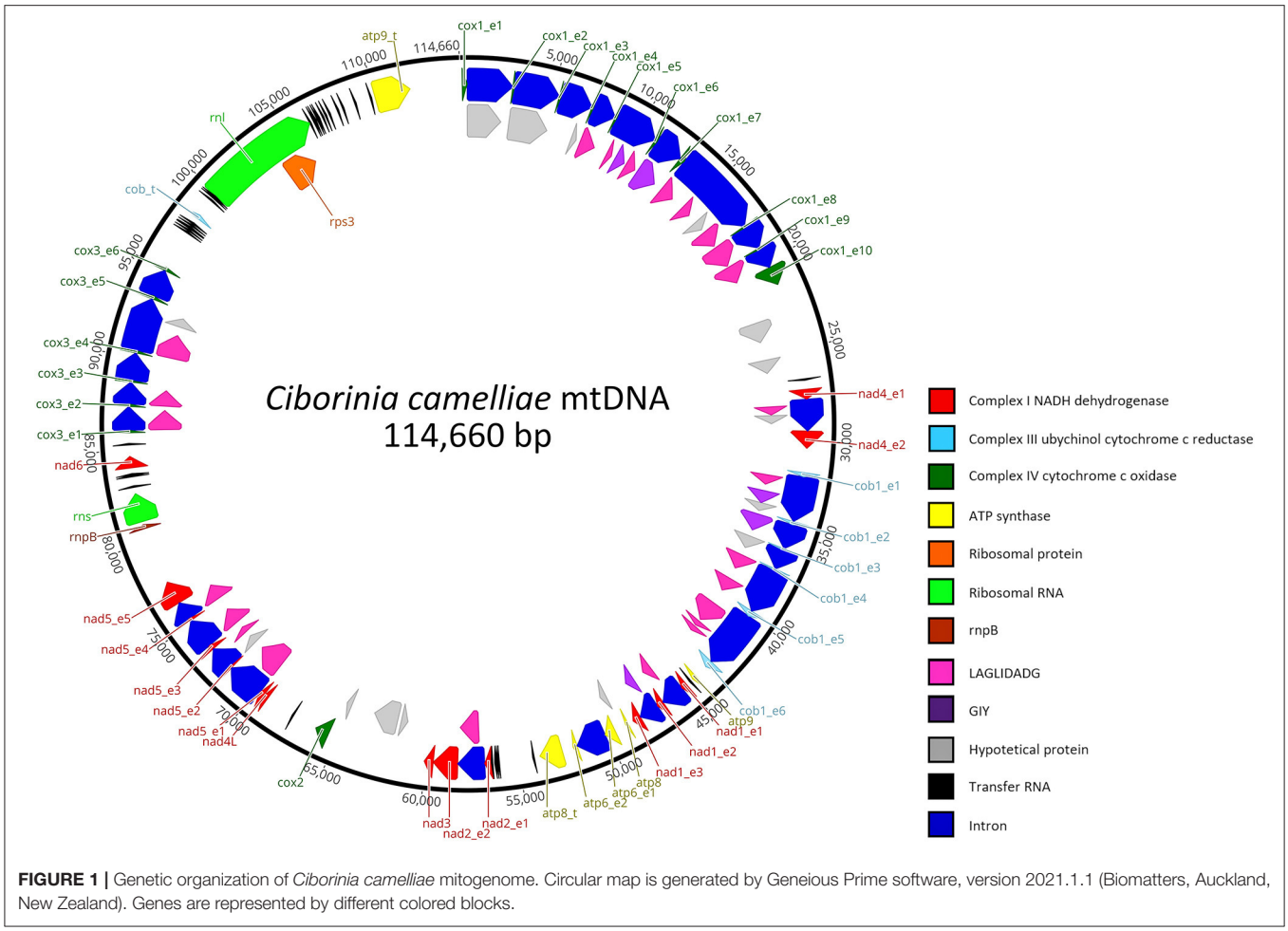


FIGURE 1 | Genetic organization of *Ciborinia camelliae* mitogenome. Circular map is generated by Geneious Prime software, version 2021.1.1 (Biomatters, Auckland, New Zealand). Genes are represented by different colored blocks.

RNA (*rnpB*) gene, and 33 transfer RNAs (tRNA) genes were detected. Overall, the RNA region accounted for 9.26% of the whole mitochondrial genome. The length of individual tRNAs ranged from 69 to 85 bp and most of these genes were placed around the *rnl* region (**Figure 1**). The total tRNAs are related to 18 essential amino acids.

In total, the mitogenome contains 62 putative genes, including 44 open reading frames (ORFs) for hypothetical proteins (15) and homing endonucleases of the LAGLIDADG (24) and GIY-YIG (5) families. BLASTp analysis was conducted for these non-conserved ORFs. The results (**Supplementary Data Sheet 1**), also verified by BLASTn analysis indicate that 22 ORFs have high identities with mitochondrial ORFs of the genus *Monilinia*. On the other hand, a high similarity with the Sclerotiniaceae species was not observed in 14 % of the ORFs. These exhibit the highest amino acid identity (ranging from 63.41 to 90.75%) with fungal species belonging to the family of Nectriaceae, Ceratocystidaceae, Cryphonectriaceae, Erysiphaceae, and Orbiliaceae. The last two are respectively evolutionary the closest and the most distant from the Sclerotiniaceae family. A LAGLIDADG of *Golovinomyces cichoracearum* and one of *Fusarium tricinctum* are the most representative ORFs with the highest amino acid identity: 85 and 90.75% with 100 and 92 % of coverage, respectively.

The coding sequences (CDS) of the 14 core mitochondrial proteins and *rps3* were used in the codon usage analysis. This investigation was carried out among the selected Sclerotiniaceae species and *F. oxysporum* as external control (**Supplementary Data Sheet 2**). The results did not show significant variability in the use of the codons. All the investigated species showed the same preferential codons, encoding the same amino acid, except for Valine and Threonine (**Table 2**). The most used codons for Threonine are ACA and ACT. In *C. camelliae* strain ITAC2, *S. sclerotiorum*, *M. laxa*, *C. shiraiana,* and *B. fuckeliana* the preferential codon is ACA, while in *M. fructicola*, *S. borealis,* and *G. lozoyensis* is ACT. The preferential codon for Valine is GTT for all investigated species, except *Monilia fructicola*, which uses more frequently GTA. Nevertheless, the GTA codon represents the second choice to encode Valine also in the other analyzed species.

Focusing on gene-specific codon usage, only in the *nad2* gene do all the investigated Sclerotiniaceae use the same preferential codons, suggesting that the *nad2* is the most conserved gene, followed by the *atp8* and *rps3* genes showing differences in only two amino acids. On the other hand, *cox1*, *atp9*, *nad3*, *cob,* and

TABLE 2 | Codon usage patterns for the 14 core mitochondrial protein coding-genes and *rps3* in Sclerotiniaceae species (*Ciborinia camelliae* strain ITAC2, *Botryotinia fuckeliana*, *Ciboria shiraiana*, *Monilinia fructicola*, *Monilinia laxa*, *Sclerotinia borealis*, *Sclerotinia sclerotiorum*) and *Glarea lozoyensis*.

	Preferential codon							
Amino acid	***Ciborinia camelliae* ITAC2**	***Botryotinia fuckeliana***	***Ciboria shiraiana***	***Monilinia fructicola***	***Monilinia laxa***	***Sclerotinia borealis***	***Sclerotinia sclerotiorum***	***Glarea lozoyensis***
Ala	GCT	GCT	GCT	GCT	GCT	GCT	GCT	GCT
Cys	TGT	TGT	TGT	TGT	TGT	TGT	TGT	TGT
Asp	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT
Glu	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA
Phe	TTT	TTT	TTT	TTT	TTT	TTT	TTT	TTT
Gly	GGT	GGT	GGT	GGT	GGT	GGT	GGT	GGT
His	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT
Ile	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA
Lys	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
Leu	TTA	TTA	TTA	TTA	TTA	TTA	TTA	TTA
Met	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG
Asn	AAT	AAT	AAT	AAT	AAT	AAT	AAT	AAT
Pro	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT
Gln	CAA	CAA	CAA	CAA	CAA	CAA	CAA	CAA
Arg	AGA	AGA	AGA	AGA	AGA	AGA	AGA	AGA
Ser	AGT	AGT	AGT	AGT	AGT	AGT	AGT	AGT
Thr	ACA	ACA	ACA	ACT	ACA	ACT	ACA	ACT
Val	GGT	GGT	GGT	GGA	GTT	GGT	GGT	GGT
Trp	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA
Tyr	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT
Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA

nad4L are the genes with the greatest differences in codons usage. They exhibited dissimilarity from 9 to 12 out of 21 amino acids (**Supplementary Data Sheet 3**).

Only for the *C. camelliae* ITAC2 strain, the codon usage of non-conserved ORFs was investigated and compared with that of conserved genes. The ORFs and the conserved regions exhibited a similar codon usage, excluding Isoleucine (ATA in conserved genes, ATT in ORFs) and Threonine amino acids (ACA and ACT in conserved regions and ORFs, respectively) (**Supplementary Data Sheet 4**).

The mitochondrial genome size is associated with a high content of introns (Deng et al., 2018). In *C. camelliae* mtDNA we detected 28 introns, which are located in eight out of 14 genes: *atp6* (1), *cob* (5), *cox1* (9), *cox3* (5), *nad1* (2), *nad2* (1), *nad4* (1) and *nad5* (4) (**Table 3**). Introns exhibit a variable length ranging from 1,015 bp to 4,750 bp. The *cox3*, *cox1,* and *cob* genes showed the highest fraction of intronic sequences: 91.2, 90.6, and 90.1%, respectively. For each gene containing introns, the exon protein-coding sequence (CDS) was <54%. In the 14 mitochondrial genes, the portion of non-coding DNA ranged from 0 to 68% (**Table 3**). The total content of non-coding DNA in the ITAC2 strain mitogenome is ∼48%. The sequences encoding the fourteen core mitochondrial proteins and *rps3* protein represent only 13.3% of the entire mitogenome.

The mtDNA of *C. camelliae* contains incomplete duplicated copies of *atp8*, *atp9,* and *cob* genes. All of these extra copies appeared truncated. The *atp8*-like ORF is located between *atp6* and *nad2* genes. The first 33 amino acids (aa) exhibit 75 % of similarity with the *atp8* protein. Instead, no significant result was found for the remaining 385 aa. The *atp9* copy is placed between *cox1* and *rnl* region. The first 59 amino acids show 100% identity with the *atp9* protein. This similarity decreases for what concerns the rest of the protein. The *cob*-like ORF was detected between *cox3* and *rnl* region and consists of only 68 aa with an identity of 80% compared to the *cob* gene. All three extra copies were observed also in *C. camelliae* ICMP 19812 strain from New Zealand (**Supplementary Data Sheet 5**). The extra genes similarity between the two *C. camelliae* strains ranges from 87.9 to 100%. The variability between the two strains was especially due to the lack of sequences or uncalled bases (Ns) in the New Zealand whole genome assembly.

A total of 36 tandem repeats were found within the *C. camelliae* mitogenome. The two longest repeat sequences measure 218 and 101 bp, respectively. The first is located in an intergenic region among tRNA genes. The second is located in the third intron of *cox1*. In total, the pair-wise nucleotide similarities ranged from 71 to 100%. Many of these regions were repeated approximately twice. The highest repetition was 7. Repetitive sequences were found also in the other investigated species (**Table 4**).

Pearson's correlation analysis was performed in species reported in **Table 4**. The association between the mitogenome

TABLE 3 | List of the 15 mt genes with their: total length, introns, protein coding sequence (CDS) size and the portion of non-coding DNA.

Gene	Gene size (bp)	n° of introns	Intronic size %	Exon CDS size (bp)	Intron CDS size (bp)	% non-coding DNA
atp6	2,436	1	68	777	0	68
atp8	144	0	0	144	0	0
atp9	222	0	0	222	0	0
cob	11,868	5	90.1	1,173	7,134	30
cox1	21,255	9	90.6	1,995	14,215	23.7
cox2	753	0	0	753	0	0
cox3	9,176	5	91.2	807	4,077	46.8
nad1	3,488	2	68.8	1,089	1,062	38.3
nad2	3,167	1	46.5	1,695	1,083	12.3
nad3	459	0	0	459	0	0
nad4	3,164	1	53.6	1,467	948	22.6
nad4L	267	0	0	267	0	0
nad5	8,052	4	75.4	1,983	4,251	18.5
nad6	672	0	0	672	0	0
rps3	1,743	0	0	1,743	0	0

Intron CDS size is the length of sequences encoding ORFs.

TABLE 4 | Comparison of *Ciborinia camelliae* strain ITAC2 mitogenome (OK326902) with closely related species: *Botryotinia fuckeliana* (KC 832409), *Ciboria shiraiana* (CM 017871.1), *Monilinia fructicola* (NC_056195.1), *Monilinia laxa* (NC_051483.1), *Sclerotinia borealis* (KJ434027), *Sclerotinia sclerotiorum* (KT283062) and *Glarea lozoyensis* (NC_031375.1) as outgroup.

Genomes features	*Ciborinia camelliae* ITAC2	*Botryotinia fuckeliana*	*Ciboria shiraiana*	*Monilinia fructicola*	*Monilinia laxa*	*Sclerotinia borealis*	*Sclerotinia sclerotiorum*	*Glarea lozoyensis*
Mitogenome size (bp)	114,660	82,212	156,608	159,648	178,357	203,051	128,852	45,501
GC content (%)	29.6	29.9	30.9	30.9	30.1	32.9	30.9	29.8
n° of conserved protein-coding genes	15	15	15	15	15	15	15	15
rRNA	2	2	2	2	2	2	2	2
tRNA	33	31	36	32	32	31	33	34
n° of introns	28	20	28	35	33	52	23	4
n° of non-conserved ORFs	44	22	77	51	94	76	30	6
Tandem repeats	36	20	27	36	60	67	27	12
n°of gene copies	3	1	1	1	0	1	1	1

size and the number of introns, non-conserved ORFs and tandem repeats was supported by statistical data (**Supplementary Data Sheet 6**). Each independent variable exhibited a high correlation coefficient with the mitogenome size, ranging from 0.929 to 0.869 (p-value < 0.05).

The arrangements of the 15 protein-coding genes were compared among some related species (**Figure 2**). Gene order in *C. camelliae* is identical to most of the Sclerotiniaceae species, except for minor differences due to replication events of *atp8*, *atp9,* and *cob* genes. Among Sclerotiniaceae, gene arrangement in *M. fructicola* appears with more evident differences. Nevertheless, all investigated species preserved four synteny units: *nad5-nad4L*, *nad3-nad2, atp6-atp8,* and *cox3-nad6*. The location of *atp9* and *cob* truncated genes of *C. camelliae* could be connected to the gene order of *Glarea lozoyensis*, even if reversed. The position of the *atp8* truncated gene between *nad2* and *atp6* was found also in *Aspergillus flavus* (NC_026920.1).

The Mauve alignment was performed only considering Sclerotiniaceae species. This investigation revealed the presence of 19 homologous regions between the seven species (**Figure 3**). The length of homologous sites ranged from 128 bp to 115,310 bp. The two largest regions contain respectively the *cox1* gene and *nad4L*, *nad5*, *rns*, *nad6,* and *cox3* gene. A homologous region (5 Kbp), containing the truncated *atp9* gene, was found only in *C. camelliae* and *S. borealis* mitogenomes.

Phylogenetic analysis was carried out using the amino acid sequences of 14 protein-coding mitochondrial genes. Nine Sclerotiniaceae species with publicly available data were investigated. To obtain a second mitochondrial set of genes from another *C. camelliae*, a partial mitochondrial

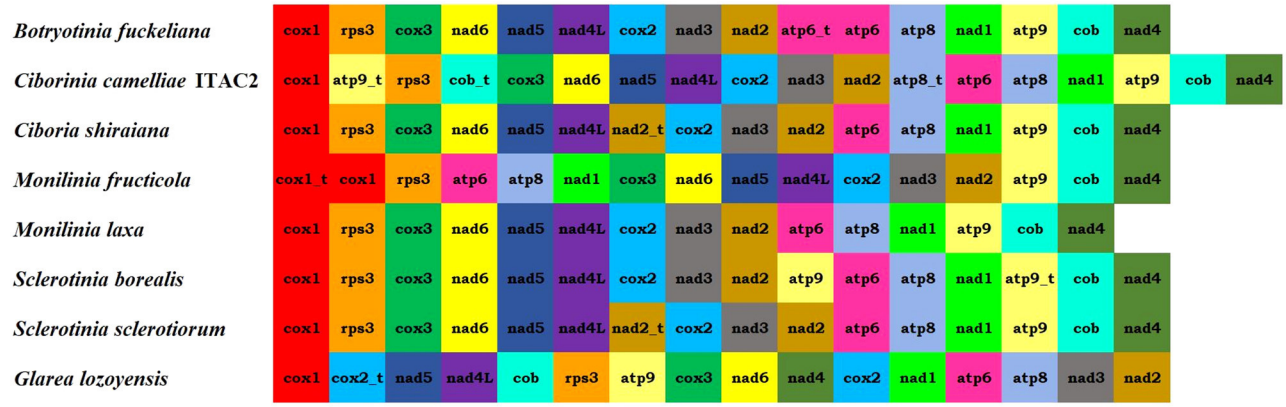


FIGURE 2 | Mitogenome organization of the seven Sclerotiniaceae investigated [*Botryotinia fuckeliana* (KC 832409), *Ciboria shiraiana* (CM 017871.1), *Ciborinia camelliae* strain ITAC2 (OK326902), *Monilinia fructicola* (NC_056195.1), *Monilinia laxa* (NC_051483.1), *Sclerotinia borealis* (KJ434027) and *Sclerotinia sclerotiorum* (KT283062)] and *Glarea lozoyensis* (NC_031375.1) as outgroup.

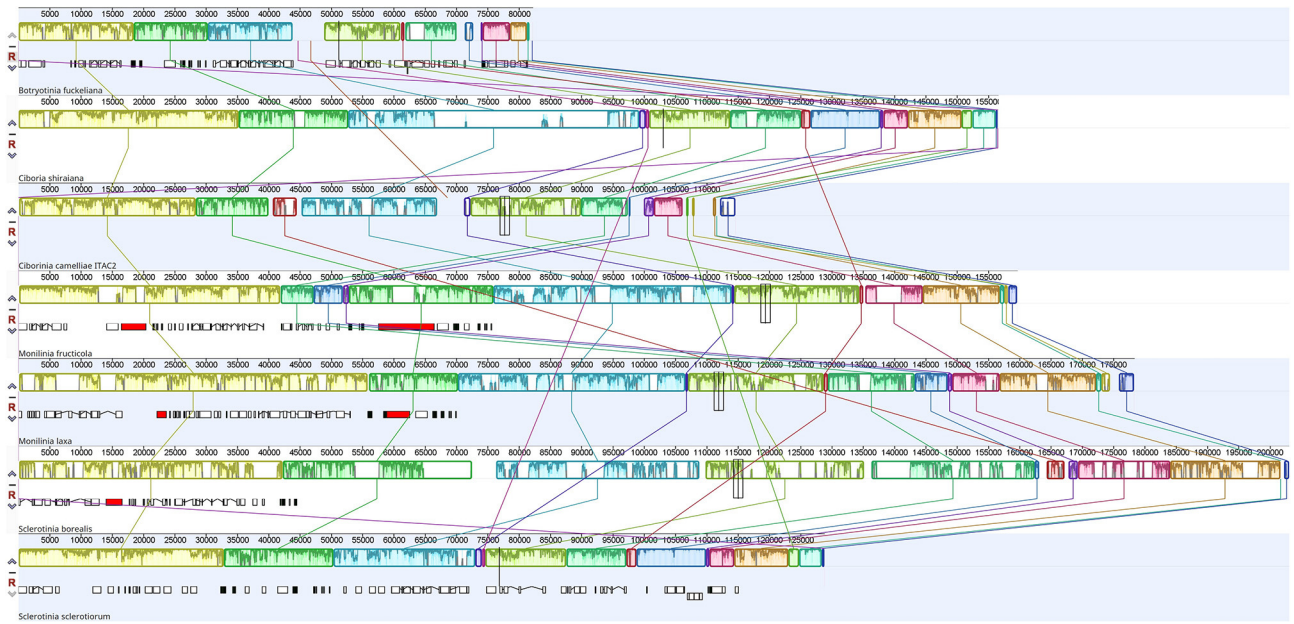


FIGURE 3 | Mauve alignment of Sclerotiniaceae mitochondrial genomes (*Botryotinia fuckeliana*, *Ciboria shiraiana, Ciborinia camelliae* strain ITAC2, *Monilinia fructicola, Monilinia laxa, Sclerotinia borealis* and *Sclerotinia sclerotiorum*). The alignment was performed setting *nad4L* gene as origin. The colored blocks reveal homologous regions between mitogenomes. Each block exhibits nucleotide similarity profile. Homologous sites are connected by lines.

genome composed of a contig list was obtained from the whole genome assembly of *C. camelliae* ICMP 19812 strain from New Zealand (**Supplementary Data Sheet 5**). *Glarea lozoyensis* was considered as an outgroup. The phylogenetic tree (**Figure 4**), executed according to the model VT+F+R3, demonstrated a close connection between *C. camelliae* ITAC2 strain and the members of the Sclerotiniaceae family. The two *C. camelliae* strains group as an independent clade, distinct from the other Sclerotiniaceae species. They exhibit 95.4 and 98.3% of similarity in nucleotide and amino acid sequence, respectively.

DISCUSSION

Nowadays, the growth of high-throughput sequencing technologies has improved phylogenetic works using mitogenomes (Zardoya, 2020). In this study, the first mitochondrial genome of *C. camelliae* was described demonstrating the phylogenetic position of the species in the Sclerotiniaceae family. The mitochondrial genome of *C. camelliae* was obtained by combining Illumina and Nanopore reads, complemented with Sanger sequencing of specific PCR products to verify the assembly. Experimentally, the combination

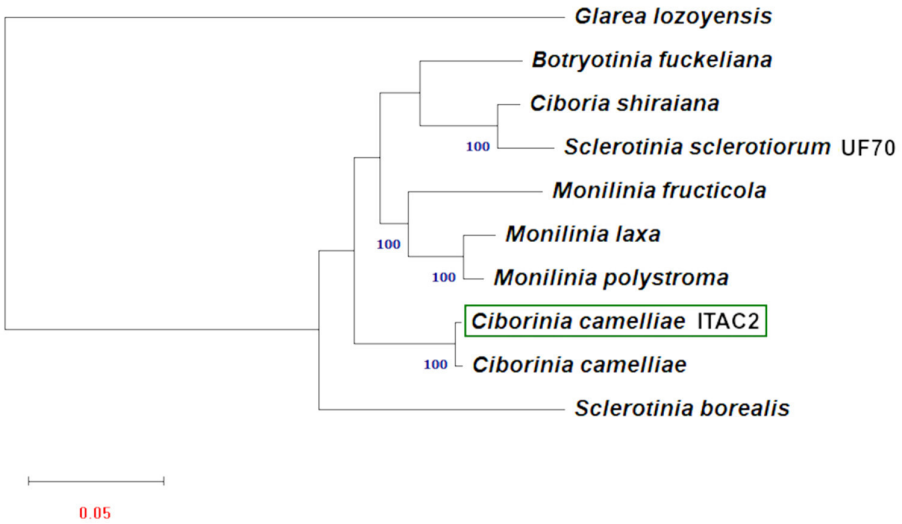


FIGURE 4 | Maximum Likelihood phylogenetic tree based on 14 encoded mitochondrial genome. Blue numbers are bootstrap support values (>90). Red number denotes branch length. The GeneBank accession numbers are listed as follow: *Botryotinia fuckeliana* (KC 832409), *Ciboria shiraiana* (CM 017871.1), *Ciborinia camelliae* (GCA_001247705.1), *Ciborinia camelliae* strain ITAC2 (OK326902) *Glarea lozoyensis* (NC_031375.1), *Monilinia fructicola* (NC_056195.1), *Monilinia laxa* (NC_051483.1), *Monilinia polystroma* (GCA_002909645.1), *Sclerotinia borealis* (KJ434027) and *Sclerotinia sclerotiorum* (KT283062).

of long and short reads allowed us to obtain a good quality mitochondrial genome as observed for other species (Degradi et al., 2021). Employing different sequencing technologies was useful to verify the existence of repetitive elements and to define their length and position in the mitogenome assembly (Kinkar et al., 2021). The Sclerotiniaceae mitogenomes were already investigated previously (Mardanov et al., 2014; Yildiz and Ozkilinc, 2020, 2021). Our work contributes to improving the knowledge on mitogenomes of the Sclerotiniaceae family by performing comparative analysis. According to the NCBI data, the *C. camelliae* strain ITAC2 mitogenome with its 114,660 bp is confirmed in the range from 82 to 203 kbp of the Sclerotiniaceae species (**Table 4**). Within this family, *B. fuckeliana* and *C. camelliae* exhibit the most compact mitochondrial genomes. Among the investigated species, *C. camelliae* mitogenome contains the highest number of truncated gene copies (*atp8*, *atp9,* and *cob*). Intermediate values in the number of non-conserved open reading frames (ORFs), introns, and tandem repeats were detected. According to Pearson's correlation analysis, the distribution of these elements, due to the evolutionary process, contributes to the high variability in mitogenome size among closely related species (Jung et al., 2010; Xiao et al., 2017; Chen et al., 2021). The intronic regions resulted the greatest promoters of variations in mitogenomes size (Li Q. et al., 2021).

Excluding the events of gene duplication, the *C. camelliae* gene arrangement is identical to most of the Sclerotiniaceae. All the examined species exhibited four syntenic units: *nad5-nad4L*, *nad3-nad2*, *atp6-atp8,* and *cox3-nad6*. These clusters were preserved also in the MAUVE alignment. One of the largest LCBs contains the *nad4L-nad5* and *nad6-cox3* pairs. The homologous regions measuring 22 and 38 Kbp include the *nad2-nad3* and *atp6-atp8* genes. These conserved clusters could have originated from a common ancestral gene order and later assumed a different organization. Some Helotiales species, such as *Phialocephala subalpine*, *Rhynchosporium orthosporum*, or *Rhynchosporium secalis*, evolutionary close to Sclerotiniaceae, exhibit the unit *nad4-nad1*, also found in some Sordariomycetes and Eurotiomycetes (Pantou et al., 2006). This cluster was instead split by *atp9* and *cob* genes in Sclerotiniaceae. In all investigated species, the *rps3* gene is located in the *rnl* region and is not interrupted by introns.

Codon usage analysis showed similar use of the optimal codons among the related species (*C. camelliae*, *B. fuckeliana*, *C. shiraiana*, *M. fructicola*, *M. laxa*, *S. borealis,* and *S. sclerotiorum*). The main differences were observed in the use of codons encoding Valine and Threonine amino acids. The same results were obtained considering less related species such as *Glarea lozoyensis* (NC_031375.1) and *Fusarium oxysporum* (NC_017930.1), suggesting the preferential use of codons is conserved among more evolutionarily distant species of fungi. Our study indicates the presence of more conserved mitochondrial regions (*nad2*, *atp8,* and *rps3*) and others prone to more variability (*cox1*, *atp9*, *nad3*, *cob,* and *nad4L*). For example, all the investigated Sclerotiniaceae showed the same codon usage in the *nad2* gene, while *atp9* and *nad3* exhibited dissimilarities in 12 amino acids. This variability may be due to weak natural selection toward synonymous substitutions not resulting in amino acid modifications (Stewart et al., 2011). Also the ORFs and the conserved regions of *C. camelliae* mtDNA showed a similar codon usage suggesting the same selective pressure for translational efficiency (Whittle et al., 2012). Nevertheless, some differences between conserved and non-conserved ORFs were observed in the differential use of synonymous codons, namely codon usage bias (CUB). The codon bias has been reported in many organisms (Duret,

2002; Sharp et al., 2005; Xiang et al., 2015; Gupta and Singh, 2021) including mitochondrial genomes (Wei et al., 2014; Yildiz and Ozkilinc, 2021). Highly-expressed genes exhibit a strong bias to conserve their translational efficiency. On the other hand, CUB decreases in non-conserved genes due to weak selection, thus they can afford to use different codons having low expression (Bulmer, 1991). On the contrary, a low CUB in weakly expressed ORFs is a mechanism for preserving a low level of expression (Grosjean and Fiers, 1982). The influence of selective pressure on the codon bias makes CUB an interesting source of species evolution force leading to their environmental adaptation (Angellotti et al., 2007; Li X. et al., 2021), which include host-pathogen relationships (Gupta and Singh, 2021). Further studies may investigate the functional role of CUB in fungal mitochondrial genomes.

As observed in other fungi (Mardanov et al., 2014; Yildiz and Ozkilinc, 2021), only the Methionine amino acid ends in G, instead, the last base of all the other preferential codons is A or T confirming the high AT content of fungal mitogenomes (Chen et al., 2019; Zhang et al., 2020).

The BLASTp analysis of the non-conserved ORFs was useful for revealing phylogenetic relationships among species. All ORFs exhibiting a significant hit in BLAST analysis, showed the best similarity with an Ascomycota species. Sixty-six percent of ORFs demonstrated the highest similarity with ORFs from the Sclerotiniaceae species. According to BLAST results, six of *C. camelliae* ORFs could have origins distinct from the Sclerotiniaceae family. Two interesting LAGLIDADG, located in *nad5* and *cox1* introns, showed 90.75 and 85% of amino acid identity with *F. tricinctum* and *G. cichoracearum*, respectively. These regions may represent events of horizontal transmission from distant fungal species, given the ability of intron I and homing endonuclease genes to move and integrate into diverse genomes (Beaudet et al., 2013; Celis et al., 2017).

The phylogenetic study based on multiple alignment of 14 concatenated mtDNA-encoded protein-related sequences further confirmed that *C. camelliae* belongs to the family of Sclerotiniaceae and its relationships among strictly related species. Phylogenetic results showed a significant difference between the two *Sclerotinia* species. Our results are partially confirmed by other works on mitochondrial DNA, where the dissimilarity between *S. borealis* and *S. sclerotiorum* was reported (Mardanov et al., 2014; Ma et al., 2019; Yildiz and Ozkilinc, 2021). The nucleotide and amino acid similarity between the two *Sclerotinia* species are 36.4 and 77.7%, respectively. Discrepancies between mitochondrial and nuclear phylogeny shall be further investigated in this genus. This phylogenetic analysis proved the closest match between the strain ITAC2 and the strain ICMP 19812 of *C. camelliae* from New Zealand, with which they constitute an independent monophyletic group. Nevertheless, some differences between the two strains at the nucleotide and amino acid sequence level could be detected in the mitochondrial genome (95.4 and 98.3% of similarity, respectively). The major level contributor to the diversity in numbers is the *cox1* gene. The Italian and New Zealand *cox1* genes exhibit 88.3% of amino acid identity, due to the deletion of some amino acids located at the beginning of *cox1* in the strain from New Zealand. According to BLAST search, the first exon of the Italian strain results lacking in the New Zealand strain. Excluding this gene, the amino acid identity between the two strains would be 99.8%. According to Mardanov et al. (2014), *cox1* is the most common target of intron insertions. This means that *cox1* represents a useful marker in phylogenetic studies, as it is evolving more rapidly than the other mitochondrial genes. The significant difference between the two *C. camelliae* strains may be due to a misassembly of the New Zealand strain genome or could represent the beginning of a speciation event. The Italian and New Zealand population are separated by geographic barriers and they may diverge in complete independence under different selection pressures (Dettman et al., 2008; Stukenbrock, 2013). Moreover, the impact of globalization on the diffusion of the phytopathogens influences the evolutionary process resulting in more variability. One day, the subsequent differentiation of populations may evolve into two different species. Large sequencing of *C. camelliae* population is warranted to further investigate the worldwide diversity of the species.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: GenBank [accession: OK326902].

AUTHOR CONTRIBUTIONS

MP, MS, and PC designed the work. IV, LD, and AK performed the experiments and acquired data. MP, MS, IV, and LD analyzed and interpreted the data for the work. IV, MP, MS, and AK drafted the manuscript. All authors revised critically the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge support from the University of Milan through the APC initiative.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/ffunb.2021.802511/full#supplementary-material

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., et al. (2014). High variability of mitochondrial gene order among fungi. *Genome Biol. Evol.* 6, 451–465. doi: 10.1093/gbe/evu028

Angellotti, M. C., Bhuiyan, S. B., Chen, G., and Wan, X. F. (2007). CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res.* 35, W132–W136. doi: 10.1093/nar/gkm392

Ballard, J. W., and Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Mol. Ecol.* 13, 729–744. doi: 10.1046/j.1365-294X.2003.02063.x

Beaudet, D., Nadimi, M., Iffis, B., and Hijri, M. (2013). Rapid mitochondrial genome evolution through invasion of mobile elements in two closely related species of arbuscular mycorrhizal fungi. *PLoS One* 8:e60768. doi: 10.1371/journal.pone.0060768

Beck, N., and Lang, B. (2010). *MFannot, Organelle Genome Annotation Websever*. Montreal, QC: Université de Montréal.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., et al. (2013). MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. doi: 10.1016/j.ympev.2012.08.023

Bolton, M. D., Thomma, B. P., and Nelson, B. D. (2006). *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. *Mol. Plant Pathol.* 7, 1–16. doi: 10.1111/j.1364-3703.2005.00316.x

Bullerwell, C. E., and Lang, B. F. (2005). Fungal evolution: the case of the vanishing mitochondrion. *Curr. Opin. Microbiol.* 8, 362–369. doi: 10.1016/j.mib.2005.06.009

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897–907. doi: 10.1093/genetics/129.3.897

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716. doi: 10.1016/j.tig.2003.10.012

Celis, J. S., Edgell, D. R., Stelbrink, B., Wibberg, D., Hauffe, T., Blom, J., et al. (2017). Evolutionary and biogeographical implications of degraded LAGLIDADG endonuclease functionality and group I intron occurrence in stony corals (*Scleractinia*) and mushroom corals (*Corallimorpharia*). *PLoS ONE* 12:e0173734. doi: 10.1371/journal.pone.0173734

Chen, C., Li, Q., Fu, R., Wang, J., Deng, G., Chen, X., et al. (2021). Comparative mitochondrial genome analysis reveals intron dynamics and gene rearrangements in two *Trametes* species. *Sci. Rep.* 11:2569. doi: 10.1038/s41598-021-82040-7

Chen, C., Li, Q., Fu, R., Wang, J., Xiong, C., Fan, Z., et al. (2019). Characterization of the mitochondrial genome of the pathogenic fungus *Scytalidium auriculariicola* (Leotiomycetes) and insights into its phylogenetics. *Sci. Rep.* 9:17447. doi: 10.1038/s41598-019-53941-5

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Degradi, L., Tava, V., Kunova, A., Cortesi, P., Saracchi, M., and Pasquali, M. (2021). Telomere to telomere genome assembly of *Fusarium musae* F31, causal agent of crown rot disease of banana. *Mol. Plant Microbe Interact.* 34, 1455–1457. doi: 10.1094/MPMI-05-21-0127-A

Deng, Y., Hsiang, T., Li, S., Lin, L., Wang, Q., Chen, Q., et al. (2018). Comparison of the mitochondrial genome sequences of six *Annulohypoxylon stygium* isolates suggests short fragment insertions as a potential factor leading to larger genomic size. *Front. Microbiol.* 9:2079. doi: 10.3389/fmicb.2018.02079

Denton-Giles, M., Bradshaw, R. E., and Dijkwel, P. P. (2013). *Ciborinia camelliae* (Sclerotiniaceae) induces variable plant resistance responses in selected species of *Camellia*. *Phytopathology* 103, 725–732. doi: 10.1094/PHYTO-11-12-0289-R

Dettman, J. R., Anderson, J. B., and Kohn, L. M. (2008). Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. *BMC Evol. Biol.* 8:35. doi: 10.1186/1471-2148-8-35

Duret, L. (2002). Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12, 640–649. doi: 10.1016/S0959-437X(02)00353-2

Galtier, N., Nabholz, B., Glémin, S., and Hurst, G. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18, 4541–4550. doi: 10.1111/j.1365-294X.2009.04380.x

Goodsell, D. S. (2010). Mitochondrion. *Biochem. Mol. Biol. Educ.* 38, 134–140. doi: 10.1002/bmb.20406

Grosjean, H., and Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 18, 199–209. doi: 10.1016/0378-1119(82)90157-3

Gupta, S., and Singh, R. (2021). Comparative study of codon usage profiles of *Zingiber officinale* and its associated fungal pathogens. *Mol. Genet. Genomics* 296, 1121–1134. doi: 10.1007/s00438-021-01808-8

Hamari, Z., Juhász, A., and Kevei, F. (2002). Role of mobile introns in mitochondrial genome diversity of fungi (a mini review). *Acta Microbiol. Immunol. Hung.* 49, 331–335. doi: 10.1556/amicr.49.2002.2-3.22

Jung, P. P., Friedrich, A., Souciet, J. L., Louis, V., Potier, S., de Montigny, J., et al. (2010). Complete mitochondrial genome sequence of the yeast *Pichia farinosa* and comparative analysis of closely related species. *Curr. Genet.* 56, 507–515. doi: 10.1007/s00294-010-0318-y

Kinkar, L., Gasser, R. B., Webster, B. L., Rollinson, D., Littlewood, D., Chang, B., et al. (2021). Nanopore sequencing resolves elusive long tandem-repeat regions in mitochondrial genomes. *Int. J. Mol. Sci.* 22:1811. doi: 10.3390/ijms22041811

Kohn, L. M. (1979). A monographic revision of the genus *Sclerotinia*. *Mycotaxon* 9, 399–400.

Kriváková, P., Cervinková, Z., Lotková, H., Kucera, O., and Rousar, T. (2005). Mitochondrie a jejich úloha v bunecném metabolismu [Mitochondria and their role in cell metabolism]. *Acta medica (Hradec Kralove)* 48, 57–67. doi: 10.14712/18059694.2018.34

Kulik, T., Brankovics, B., van Diepeningen, A. D., Bilska, K., Zelechowski, M., Myszczyński, K., et al. (2020). Diversity of mobile genetic elements in the mitogenomes of closely related *Fusarium culmorum* and *F. graminearum* sensu stricto strains and its implication for diagnostic purposes. *Front. Microbiol.* 11:1002. doi: 10.3389/fmicb.2020.01002

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, Q., Li, L., Feng, H., Tu, W., Bao, Z., Xiong, C., et al. (2021). Characterization of the complete mitochondrial genome of basidiomycete yeast *Hannaella oryzae*: intron evolution, gene rearrangement, and its phylogeny. *Front. Microbiol.* 12:646567. doi: 10.3389/fmicb.2021.646567

Li, X., Wang, X., Gong, P., Zhang, N., Zhang, X., and Li, J. (2021). Analysis of codon usage patterns in *Giardia duodenalis* based on transcriptome data from *Giardia*DB. *Genes* 12:1169. doi: 10.3390/genes12081169

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M. W., Chaisson, M., and Pevzner, P. A. (2016). Assembly of long error-prone reads using de Bruijn graphs. *Proc. Natl. Acad. Sci.* 113, E8396–E8405. doi: 10.1073/pnas.1604560113

Lv, J., Bhatia, M., and Wang, X. (2017). Roles of mitochondrial DNA in energy metabolism. *Adv. Exp. Med. Biol.* 1038, 71–83. doi: 10.1007/978-981-10-6674-0_6

Ma, Y., Huang, L., Abuduaini, A., Zhou, H., Wang, Y., and Suo, F. (2019). Complete mitochondrial genome of plant pathogen *Monilinia fructicola* (Sclerotiniaceae, Helotiales). *Mitochondrial DNA B.* 4, 791–792. doi: 10.1080/23802359.2019.1567282

Mardanov, A. V., Beletsky, A. V., Kadnikov, V. V., Ignatov, A. N., and Ravin, N. V. (2014). The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. *PLoS ONE* 9:e107536. doi: 10.1371/journal.pone.0107536

Medina, R., Franco, M., Bartel, L. C., Martinez Alcántara, V., Saparrat, M., and Balatti, P. A. (2020). Fungal mitogenomes: relevant features to planning plant disease management. *Front. Microbiol.* 11:978. doi: 10.3389/fmicb.2020.00978

Newmeyer, D. D., and Ferguson-Miller, S. (2003). Mitochondria: releasing power for life and unleashing the machineries of death. *Cell* 112, 481–490. doi: 10.1016/S0092-8674(03)00116-8

Pantou, M. P., Kouvelis, V. N., and Typas, M. A. (2006). The complete mitochondrial genome of the vascular wilt fungus *Verticillium dahliae*: a novel gene order for *Verticillium* and a diagnostic tool for species identification. *Curr. Genet*. 50, 125–136. doi: 10.1007/s00294-006-0079-9

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

Saracchi, M., Colombo, E. M., Locati, D., Valenti, I., Corneo, A., Cortesi, P., et al. (2022). Morphotypes of *Ciborinia camelliae Kohn* infecting camellias in Italy. *J. Plant Pathol.* 104. doi: 10.1007/s42161-022-01040-2

Saracchi, S., Locati, D., Colombo, E. M., and Pasquali, M. (2019). Updates on *Ciborinia camelliae*, the causal agent of camellia flower blight. *J. Plant Pathol*. doi: 10.1007/s42161-018-0173-0

Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F., and Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33, 1141–1153. doi: 10.1093/nar/gki242

Shingu-Vazquez, M., and Traven, A. (2011). Mitochondria and fungal pathogenesis: drug tolerance, virulence, and potential for antifungal therapy. *Eukaryotic Cell* 10, 1376–1383. doi: 10.1128/EC.05184-11

Stewart, J. E., Kawabe, M., Abdo, Z., Arie, T., and Peever, T. L. (2011). Contrasting codon usage patterns and purifying selection at the mating locus in putatively asexual *Alternaria* fungal species. *PLoS ONE* 6:e20083. doi: 10.1371/journal.pone.0020083

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* 28, 1102–1104. doi: 10.2144/00286ir01

Stukenbrock, E. H. (2013). Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytol.* 199, 895–907. doi: 10.1111/nph.12374

Taylor, C. H., and Long, P. G. (2000). Review of literature on camellia flower blight caused by *Ciborinia camelliae*. *N. Z. J. Crop Horticult. Sci*. 28, 123–138. doi: 10.1080/01140671.2000.9514132

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 44, W232–W235. doi: 10.1093/nar/gkw256

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Res*. 40:e115. doi: 10.1093/nar/gks596

van Toor, R. F., Ridgway, H. J., Butler, R. C., Jaspers, M. V., and Stewart, A. (2005). Assessment of genetic diversity in isolates of *Ciborinia camelliae* Kohn from New Zealand and the United States of America. *Australas Plant Path* 34, 319–325. doi: 10.1071/AP05040

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963

Wei, L., He, J., Jia, X., Qi, Q., Liang, Z., Zheng, H., et al. (2014). Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evol. Biol*. 14:262. doi: 10.1186/s12862-014-0262-4

Whittle, C. A., Sun, Y., and Johannesson, H. (2012). Genome-wide selection on codon usage at the population level in the fungal model organism *Neurospora crassa*. *Mol. Biol. Evol.* 29, 1975–1986. doi: 10.1093/molbev/mss065

Xiang, H., Zhang, R., Butler, R. R. 3rd, Liu, T., Zhang, L., Pombert, J. F., and Zhou, Z. (2015). Comparative analysis of codon usage bias patterns in microsporidian genomes. *PLoS ONE* 10:e0129223. doi: 10.1371/journal.pone.0129223

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

Yildiz, G., and Ozkilinc, H. (2020). First characterization of the complete mitochondrial genome of fungal plant-pathogen *Monilinia laxa* which represents the mobile intron rich structure. *Sci. Rep*. 10:13644. doi: 10.1038/s41598-020-70611-z

Yildiz, G., and Ozkilinc, H. (2021). Pan-mitogenomics approach discovers diversity and dynamism in the prominent brown rot fungal pathogens. *Front. Microbiol*. 12:647989. doi: 10.3389/fmicb.2021.647989

Zardoya, R. (2020). Recent advances in understanding mitochondrial genome diversity. *F1000Research* 9, 1–19. doi: 10.12688/f1000research.21490.1

Zhang, Y., Yang, G., Fang, M., Deng, C., Zhang, K. Q., Yu, Z., et al. (2020). Comparative analyses of mitochondrial genomes provide evolutionary insights into nematode-trapping fungi. *Front. Microbiol.* 11:617. doi: 10.3389/fmicb.2020.00617

Conflict of Interest: The 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.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fmicb.2022.866187

Refining Mitochondrial Intron Classification With ERPIN: Identification Based on Conservation of Sequence Plus Secondary Structure Motifs

*Samuel Prince†, Carl Munoz†, Fannie Filion-Bienvenue, Pierre Rioux, Matt Sarrasin and B. Franz Lang**

Département de Biochimie, Robert Cedergren Center for Bioinformatics and Genomics, Université de Montréal, Montréal, QC, Canada

OPEN ACCESS

Edited by:
Vassili N. Kouvelis, National and Kapodistrian University of Athens, Greece

Reviewed by:
Georg Hausner, University of Manitoba, Canada
Weilong Hao, Wayne State University, United States

***Correspondence:**
B. Franz Lang
b.franz.lang@umontreal.ca

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 January 2022*
Accepted: *28 February 2022*
Published: *18 March 2022*

Citation:
Prince S, Munoz C, Filion-Bienvenue F, Rioux P, Sarrasin M and Lang BF (2022) Refining Mitochondrial Intron Classification With ERPIN: Identification Based on Conservation of Sequence Plus Secondary Structure Motifs. Front. Microbiol. 13:866187. doi: 10.3389/fmicb.2022.866187

Mitochondrial genomes—in particular those of fungi—often encode genes with a large number of Group I and Group II introns that are conserved at both the sequence and the RNA structure level. They provide a rich resource for the investigation of intron and gene structure, self- and protein-guided splicing mechanisms, and intron evolution. Yet, the degree of sequence conservation of introns is limited, and the primary sequence differs considerably among the distinct intron sub-groups. It makes intron identification, classification, structural modeling, and the inference of gene models a most challenging and error-prone task—frequently passed on to an "expert" for manual intervention. To reduce the need for manual curation of intron structures and mitochondrial gene models, computational methods using ERPIN sequence profiles were initially developed in 2007. Here we present a refinement of search models and alignments using the now abundant publicly available fungal mtDNA sequences. In addition, we have tested in how far members of the originally proposed sub-groups are clearly distinguished and validated by our computational approach. We confirm clearly distinct mitochondrial Group I sub-groups IA1, IA3, IB3, IC1, IC2, and ID. Yet, IB1, IB2, and IB4 ERPIN models are overlapping substantially in predictions, and are therefore combined and reported as IB. We have further explored the conversion of our ERPIN profiles into covariance models (CM). Current limitations and prospects of the CM approach will be discussed.

Keywords: mitochondrial introns, group I, ERPIN, covariance models, infernal, RNA structure

INTRODUCTION

Sequencing of mitochondrial (mt) genomes (separately or as part of whole-genome projects) has become easy and affordable, but identifying and annotating genes in mt contigs often remains challenging. This is because mt genes, particularly in fungi, may contain a substantial number of (sometimes large) Group I and Group II introns, as well as difficult-to-recognize

mini-exons that can be as small as three (Cummings et al., 1990b) or, at an extreme, a single nucleotide (Osigus et al., 2017).

In nuclear genome projects, the inference of gene models can leverage transcript alignments, in conjunction with alignments of conserved protein or structured RNA sequences from related species onto the genome sequence. However, mitochondrial transcript data are not only rarely available but also of limited help, as splicing of mt RNA precursors tends to be partial and is often difficult to interpret without expert manual intervention. Therefore, mitochondrial gene model inferences are usually only based on the set of conserved mitochondrial gene or derived protein sequences (Paquin et al., 1997; Gray et al., 1999; Lang et al., 1999; Burger et al., 2003; Lang, 2013). Evidently, this approach has serious drawbacks. When relying on publicly available sequences, these must be closely related to the genome to be annotated, and *a priori* be complete and accurate, otherwise implicit errors will occur *via* "community error propagation." It is likewise possible to curate the gene annotations of neighboring species case by case, an approach that requires substantial input of a knowledgeable expert curator. Moreover, sequence matching of known coding or protein sequences (which is employed in both aforementioned approaches) can be fairly precise for delineation of larger exons, but can often fail for those smaller than ~30 nt, particularly when two or more small exons are "hiding" in long stretches of intron-rich sequence. It is here that high confidence and complete intron identification plays a crucial complementary role in revealing approximate locations of potential exons (i.e., in stretches of sequence between predicted introns). In addition, structural RNA inference of introns can provide clues on precise exon–intron boundaries flanked by conserved sequence features.

In the following we will first explain the distribution and general structural features of these introns, with emphasis on mt Group I and its sub-groups. Group II introns will not be further discussed as they are readily identified computationally, with a single search (Lang et al., 2007), based on two small adjacent helical regions (domains V plus VI). In stark contrast, Group I intron identification works very poorly with a general intron model and instead requires searching with models that represent the distinct sub-groups as well as structurally derived intron variants (Lang et al., 2007). We will then go on to explain two powerful search algorithms [ERPIN (Gautheret and Lambert, 2001) and Infernal (Nawrocki et al., 2009)] that are best suited for identifying these structured RNAs and their sub-groups, weighing advantages and potential drawbacks.

Distribution and Structural Features of Group I and II Introns

Group I and II introns occur in a wide range of mitochondrial, chloroplast, eubacterial, bacteriophage, virus, and nuclear genomes [e.g., (Michel et al., 1982; Cech et al., 1983; Michel and Dujon, 1983; Shub et al., 1988; Ferat and Michel, 1993; Ohta et al., 1993; Turmel et al., 1993; Michel and Ferat, 1995; Qiu et al., 1998; Cho and Palmer, 1999). They are (or originated from) mobile elements that spread *via* intron-encoded proteins most notably "homing" endonucleases in Group I (Dujon, 1989; Henke et al., 1995; Belfort and Bonocora, 2014), and reverse transcriptases in Group II (Michel and Lang, 1985; Schäfer et al., 2003; Lambowitz and Zimmerly, 2004)]. In contrast to the eukaryotic spliceosomal introns in nuclear genes, Group I and II introns are characterized by elaborate, conserved (but unrelated) RNA structures that were first recognized in the early 1980s in fungal mtDNAs [e.g., (Dujon, 1980; Michel et al., 1982; Waring et al., 1982)]. Group I introns were shortly thereafter identified in ciliate nuclear rRNA genes and were demonstrated to self-splice *in vitro* without requiring protein factors (Kruger et al., 1982; Cech et al., 1983). This finding motivated a large number of research groups to investigate the "self-splicing" properties of Group I and II introns that were identified in their sequencing projects, to rather mixed results. Successful *in vitro* splicing in the absence of protein co-factors was reported for only a limited number of introns. More often, splicing depends on intron-encoded proteins [termed "maturases," e.g., (Carignani et al., 1983)], or on proteins encoded in separate nuclear genes [e.g., (Kreike et al., 1987; Augustin et al., 1990; Bassi et al., 2002; Huang et al., 2005)]. In particular, mitochondrial introns turned out to be frequently unable to splice *in vitro* in the absence of protein co-factors [e.g., (Schäfer et al., 1991)], even under most un-physiological test conditions, like high salt, temperature, etc. (and, such negative results are evidently under-reported in the literature). Accordingly, the general notion that Group I and II introns are autocatalytic or self-splicing is quite misleading. Qualifying them as ribozymes, which in some instances undergo autocatalytic splicing *in vitro*, appears to be more in line with the published literature.

Mt. Group I introns were initially classified into Group IA, IB, IC, and ID [with an additional bacterial IE group added a few years later; for a review see (Hausner et al., 2014)], and further subdivided (e.g., IC1 and IC2). The initial mt group II intron subdivisions are Group IIA and IIB, later extended with the identification of a bacterial IIC (Zimmerly and Semper, 2015). Although reaching back as much as 32 years, this classification continues to be widely used and accepted. Group I classification is still based on the 87 available sequences at the time [see appendix in (Michel and Westhof, 1990)], collected from organelle and bacteriophage genomes, plus introns in ciliate nuclear rRNA genes. Notably, the vast majority of sequences came from fungal mt genomes, with more than one-third (38%) from a single species, *Podospora anserina* (Cummings et al., 1990a; Michel and Westhof, 1990). Evidently, this sampling is highly biased toward mt introns, and any of these groupings rely essentially on human expertise, rather than on computational methods. In the absence of a sufficient number of intron sequences per sub-group, which would have allowed a phylogeny- or sequence profile-based grouping, the initial ordering of group I introns into sub-groups gave most (but not all) weight to the P7 pairing, which is an essential part of the catalytic core of the ribozyme serving as a binding site for a guanosine cofactor (Michel et al., 1989). Other relatively well-conserved regions that were considered in addition are the P1 stem that defines the 5′ splice site and the P4-P5-P6

and P3-P7-P9 helices (Michel and Westhof, 1990). Note that substantial variation of the P7 motif was accepted within given sub-groups as long as overall structural or sequence relatedness was recognized, which speaks against the popular characterization of intron groups *via* P7 sequence motifs ("logos"). In fact, logos emphasize the most predominant primary sequence, therefore lack detail on sequence and structural variation (i.e., the characteristic helix-bulge-helix of P7) that is essential for ribozyme catalysis.

Computational Methods for Intron Identification

Basic similarity search algorithms, as implemented in BLAST and FASTA are woefully inadequate in identifying introns for two reasons. The first being relatively high levels of sequence variation in introns, which can degrade the quality of high-scoring sequence pairs, and thus lead to imprecise and fragmented hits. The second reason is that similarity comparisons are blind to secondary structure, which limits their capacity to bridge distant conserved motifs. Instead, probabilistic approaches using sequence profiles (based on structured alignments of multiple sequences, including information of secondary structure pairings) are required to spot the regions of similarity that are small and spread out over intron sequences that can reach up to 7 kb [e.g., (Liu et al., 2020)] and beyond. Currently available and popular search algorithms are ERPIN (Gautheret and Lambert, 2001) and Infernal (Nawrocki et al., 2009). ERPIN is based on column-wise computation of probabilities at the nucleotide and structure level, focusing on the detection of distinct conserved sequence motifs and helices in given structured sequence alignments (to be supplied by the user). In contrast, Infernal leverages the HMM approach, computing emission (at a given column) and transition probabilities (from one column to the next), but applies covariance modeling (CM) as a second layer search mechanism to initial HMM hits. The CM architecture is a stochastic context-free grammar (SCFG) profile which, in the same spirit of HMMs, consists of states (with emission and transition probabilities) associated with the single nucleotides and pairs that make up the RNA structure. CMs are therefore expected to be more sensitive than ERPIN, and because of the underlying HMM approach that in contrast to ERPIN allows for insertions and deletions that are not identified as such in the search model, useful in improving structured sequence alignments.

In 2006, 16 years after the initial Group I intron classification by Michel and Westhof, the increased number of available organelle genomes (then the most substantial and diverse source of intron sequences) allowed the development of intron search models for automatic identification and classification of virtually all known organelle group I and II introns with high confidence (Lang et al., 2007). The underlying algorithm for this approach has been **ERPIN** (Gautheret and Lambert, 2001). Yet even in 2006, the low number of sequences in some intron sub-groups had limited automated approaches with ERPIN because a computationally objective confirmation of intron group consistency was out of reach. As a consequence, structured sequence alignments may have in rare instances included a sequence from an unrelated sub-group, potentially leading to intron predictions in both the target and contaminant sub-groups. Notwithstanding, the use of these ERPIN search models has been reasonably precise and complete (Lang et al., 2007), which was a requirement for developing our MFannot mitogenome annotator.[1]

Since 2006, no systematic update of our intron sequence alignments has been conducted to verify the ERPIN approach and the findings. At the algorithmic level however, the development of **covariance models** (**CM**; Nawrocki et al., 2009) have become an attractive alternative to ERPIN, due to a recent substantial performance increase (Nawrocki et al., 2009; Nawrocki and Eddy, 2013), resulting in search times comparable to, if not better than, ERPIN. In fact, CM analysis has enabled detection of the widespread presence of group IA3 and IB4 introns in Archaea (Nawrocki et al., 2018). The CM approach has not yet been compared against ERPIN, or more broadly verified for both its sensitivity and its precision in sub-group classification. Incidentally, a recent study had leveraged CMs uniquely in the context of mitochondrial group ID introns, limited in scope to both the core motifs as well as to the relatively narrow lineage of pezizomycete fungi (Cinget and Bélanger, 2020). Furthermore, the aligned ID intron sequences were taken from the now defunct GISSD intron database (Zhou et al., 2008), which implies the quality of the underlying data must be taken for granted. Thus, the specificity and sensitivity of the resulting CM to the ID group remain to be clarified.

Challenges in Assembling a Consistent Set of Group I Intron Predictors

The currently available approaches for modeling RNA sequences with 2D structure layered on, ERPIN and Infernal/CM, have both specific advantages and drawbacks. The strength of ERPIN is in examining clearly defined structural or sequence motifs, by providing the user with the option of identifying distinct motifs and searching them in any given combination and order. The ERPIN search strategy can be optimized to be both sensitive and rapid in execution (despite lack of parallelization), by searching highly conserved motifs at an initial level (preferentially single-stranded region that are much more rapidly identified than helical interactions), and followed by inclusion of other peripheral motifs. It is important to realize that this motif-driven approach allows for modeling of pseudoknots, which in Group I introns include the universal P7 structural motif—a crucial element of the ribozyme catalytic domain (Michel and Westhof, 1990). ERPIN requires that conserved motifs, such as P7, be supplied together with a structural multiple sequence alignment, which can often be a challenging task. Another clear drawback of ERPIN vs. Infernal/CM is its unforgiving rigidity in defining a search model with distinct sequence or secondary structure motifs. For instance, nucleotide deletions in helical regions of search models are not allowed, contrary to the CM approach, which also accepts and then properly

[1]https://megasun.bch.umontreal.ca/RNAweasel/

aligns nucleotide insertions and deletions (indels) of the resulting hits. In addition, partial hits will not be reported by ERPIN, which is an issue with derived intron structures that carry shortened or completely lacking motifs. Finally, too much sequence variation of a target ncRNA may result in ERPIN models that produce few or even no results. A solution proposed by the authors is a subdivision of sequences and respective ERPIN models, a "divide and conquer" strategy that we already successfully employed with our initial set of ERPIN intron predictors (Lang et al., 2007).

Infernal (cmsearch) on the other hand does report partial hits, and has substantially better sensitivity in sequence motif identification, as it uses an **HMM-SCFG approach** of assigning emission and transition probabilities (rather than the column-wise probabilities of ERPIN). This may be relevant as the current implementation of CMs proscribe strict processing from 5′ to 3′ of the given model, thus treats pseudoknots only at the primary sequence level. Yet, as long as the pseudoknot motif has significant nucleotide sequence conservation (which is not necessarily the case for Group I introns), the increased sensitivity of the HMM approach may (or not) compensate for the lack of pseudoknot helix modeling. Additionally, and in contrast to EPRIN, cmsearch excels at finding matches with CMs containing only few aligned sequences. The execution times of cmsearch may be faster than equivalent ERPIN searches, depending largely on given search models and the available CPU, as cmsearch natively supports parallelized computations. Taken together, CM alignments are attractive for the expert development of alignments because of its flexibility in finding matches and because of the formatting of results as a structured alignment. To be clear however, this functionality does not liberate the user from providing an initial multiple sequence alignment together with a 2D structure line. For this, a pre-alignment at the sequence level with one of the many multiple sequence alignment tools [e.g., Muscle (Edgar, 2004)], followed by prediction of secondary structure pairings [e.g., RNAalifold (Bernhart et al., 2008) or R-scape (Rivas et al., 2020)] will provide a structured alignment that still needs to be refined by an expert.

Short Term and Long-Term Objectives

The unprecedented number of mt genomes that have been added more recently to the GenBank repository has progressed from a severe lack of sequence data to "land of milk and honey" with regard to intron analysis. In this paper, we will focus on the 662 fungal mitochondrial (mt) DNAs, identified in various sections of GenBank by November 2021 (see below), because of their most elevated intron density [e.g., 81 in *Endoconidiophora resinifera*, (Zubaer et al., 2018)], broadly covering all but the more recently identified nuclear and bacteria-specific sub-groups. Our objective for intron model building is automated alignment of well-conserved and universally present motifs in currently defined intron sub-groups, starting as a test case with mt Group I introns [i.e., as originally defined in the seminal Michel and Westhof publication (Michel and Westhof, 1990)]. The resulting structural models will be tested for overlapping predictions, either for dismissal of traditional sub-groups or the inference of additional ones. The questions that we will address in the context of intron identification and classification are as follows:

- is the currently accepted intron sub-grouping for IA, IB, IC, and ID valid and consistent from an evolutionary/computationally point of view;
- does automated, probabilistic intron classification with ERPIN identify known fungal mt introns within the given sub-groups and without ambiguity.

We will conclude with a brief preview on covariance searches with Infernal/CM, to test whether CMs are as performant in intron identification, and as suitable for intron sub-grouping as ERPIN.

MATERIALS AND METHODS

Building of ERPIN Intron Models

As a starting point, aligned intron sequences for each sub-group that are listed in the Michel and Westhof publication (Michel and Westhof, 1990) were shortened to the virtually ubiquitous helical regions P2,4,6,7,8 and adjacent conserved single-stranded sequences, and a respective 2D structure line in ERPIN format was added to the sequence alignments. The resulting models were searched against the set of 662 fungal mtDNAs using the ERPIN (version 5.5.4) wrapper script RNAweasel that allows to eliminate identical as well as closely related sequences in the aligned training set. The aligned matches were added to the current ERPIN model to create an extended ERPIN model, visually inspected for potential misalignments, manually corrected or adapted (i.e., by extending the size of variable insert regions that separate the conserved sequence motifs), or otherwise discarded if inconsistent with the overall alignment. After repeating this process several times, the consistency of the resulting model was tested by searching against the Michel collection of sequences, expecting to match only members within the same intron group, or at most identifying conflicting matches with low scores. The same type of test was conducted with sequences from the GISSD database, and finally, the results of all our searches with different ERPIN models were checked for conflicting matches.

Identification of Conflicting Matches With Different ERPIN Models

Conflicting matches were identified with a Python script that analyses the coordinates of hits across multiple RNAweasel log files, to flag intron intervals that are predicted by more than one ERPIN model. First, conflicting hits (i.e., sharing at least 1 nt between their corresponding genomic intervals) were assigned to a group. For each group, the proportion of shared conserved nucleotides (in capital letters in the log file) between the hits was computed to aid in the separation of the hits into the two categories "conflicting prediction" or "overlapping introns." The final parsing of the result was done manually; hits that shared the same (or almost the same) start and end position were labeled "conflicting" (>95% identical conserved positions found by both models) while the rest of the hits

were identified as "overlapping introns." For the IB sub-group analysis, the parsing was done automatically without distinction between "conflicting" and "overlapping introns."

Development of ERPIN Search Strategies

Finding an optimal search strategy for every given ERPRIN model is essential for execution speed, sensitivity of searches as well as appropriate cutoff values. According to previous experience search strategies with three (rarely four) search levels are most effective (using-add statements as described in the ERPIN manual). The initial search level will pinpoint potentially intron-containing genomic regions, with subsequent search levels selecting those that meet the full set of constraints (for more details on the principals of element regrouping and order of search levels, see the main text). Initially, Skylign (Wheeler et al., 2014) is used to generate a logo of a multiple sequence alignment (MSA). Skylign converts the MSA to a Hidden Markov Model (HMM) in order to estimate position-specific (including gaps) probability distributions, or logo stack heights. The letter proportions per stack (or, position) are computed from the respective estimated nucleotide probabilities. Once the logos are obtained, an in-house script computes the average stack height across each distinct motif (e.g., segment of single-strand, or segment half of a paired sequence). Motifs with higher average probabilities are identified and labeled according to position on the secondary structure line. Motifs with lower average probabilities are subsequently defined. Finally, the script combines the motif definitions, along with cutoff scores derived from the (ERPIN) tstat summaries of the respective motifs, to automatically create optimized parameters for an intron model search.

Conversion of ERPIN Model Alignments Into CM-Compatible Stockholm Format

ERPIN models have a custom-encoding of the structure line information (based on consecutive numbering of elements; if a number occurs only once, it is a single-stranded region; if the number occurs twice, it defines a helical interaction). In order to permit the use of the ERPIN model information for building and calibration of respective CMs, the ERPIN format needs to be converted into Stockholm format, which uses bracket expressions for identification of helices, but requires additional encoding with pairs of upper/lower case characters (known as WUSS shorthand) to identify the pseudoknot that is present in most ERPIN intron models. Structure lines containing a pseudoknot, and modified by converting the helical components to WUSS shorthand (AA..aa), will be recognized by Infernal tools and interpreted only at the nucleotide conservation level.

RESULTS AND DISCUSSION

Collecting and Formatting All Publicly Available Fungal mtDNAs—A Non-trivial Task

The success of our project to update Group I and II intron search models critically relies on the availability of a taxonomically broad and complete collection of fungal mtDNA sequences. The National Center for Biotechnology Information (NCBI) sequence databases have continued to grow exponentially, including the Organelle Genome Database (OGD) that housed as many as 12,582 mitochondrial genome assemblies as of October 19th, 2021. While the OGD appears to be the NCBI's front-end to searching mitochondrial genome records, assemblies can also be found in the NCBI Genome Reports (GR) section and in the continually updated RefSeq release of mitochondrial records. Searches against the Nucleotide (NT) database reveal again additional mitochondrial assemblies of varying quality and completeness. There are several issues with the vast databases of NCBI that hamstring studies aimed broadly at fungi. The first issue is that the fungal lineage represents only a minority of records across all databases, as shown in **Supplementary Table S1**. The number of fungal records typically make up less than 5% of the total, with the exception of the NCBI Genome Reports, after taking into account redundant accessions. The second issue is related to inconsistencies in accessions listed across the above-mentioned databases. As shown in **Supplementary Figure S1**, accessions listed under the GR appear to be completely contained within the OGD, but there is a handful of records uniquely found in RefSeq, and more than 1700 unique to NT, which can introduce a large sample size bias. The last issue is related to gaps in the GenBank records themselves. Full taxonomic, annotation, and sequence information is typically bundled in GenBank records, but some records, mostly from the OGD and GR, are missing the underlying genomic sequence. Furthermore, almost all of the incomplete records are the same in both the OGD and the GR, whereas a minority are incomplete in the curated RefSeq collection (**Supplementary Figure S2**). Ironically, the largest database of GenBank records, NT, has the least number of incomplete records.

In order to assemble a taxonomically broad collection of fungal mt genomes, we extracted records from the OGD[2] and GR (see Footnote 2) front-ends and then combined into a preliminary list. The NT collection (in compressed fasta format) was then downloaded[3] and filtered for keywords related to partial or complete mitochondrial genome assemblies. The resulting accessions are parsed and added to the OGD and GR collection. A python script, leveraging the BioPython modules (Cock et al., 2009), was written to parse GenBank files and output a clean fasta nucleotide sequence record for each mitochondrial contig (i.e., disregarding information on gene, exon, and intron positions), with a header formatted to include a short unique ID (derived from genus, species, and accession) in the first field, followed by full genus and species name, then full taxonomic information, and capped with the original accession. Such a format is more conducive to careful selection of species diversity, and visual inspection of phylogenetic trees given how current phylogeny tools function.

[2]https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/fungi
[3]https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz

Automation of ERPIN Search Strategies, and Phylogenetic Filtering of Structured Alignments

Since searches involving multiple motifs may quickly become demanding if not unfeasible in terms of CPU time and memory usage, ERPIN provides the option of a **multi-leveled search strategy** (Gautheret and Lambert, 2001; Lambert et al., 2004). It allows grouping of (sequence or structural) motifs that are searched in iterative steps. The initial search level will pinpoint potentially intron-containing genomic regions, with subsequent search levels selecting those that meet the full set of constraints. In other words, the execution speed depends essentially on the choice of the first-level motifs—preferentially well-conserved single-stranded regions that are most rapidly identified. Yet, finding the best search strategy requires a rather lengthy trial and error optimization by the model developer—motivating the development of a more objective computational procedure (Auto-strategy; in-house script available on request). It takes an ERPIN model (i.e., a set of aligned sequences with secondary structure predictions) and a collection of targeted genome sequences as an input and constructs a search strategy with corresponding cutoff values for the given ERPIN model and sequence collection. More specifically, the algorithm (for details see Methods) allows the computation of a three- or four-level search following several principles: (1) selection of closely spaced and strongly conserved sequence motifs for level one, to enable speedy initial searches; (2) at all levels, combination of several conserved motifs, sufficient to avoid false positives; and (3) a final addition of the remaining elements. Auto-strategy often results in more effective and specific searches compared to manual strategies and may serve as a quickly computed starting point for further optimization by the model developer. The tests carried out during the development of Auto-strategy were performed on sub-groups A1, A3, B, C1, C2, and D of mitochondrial group I introns. The number of hits obtained by the automatic strategies was frequently higher compared to the manual ones. Yet, subsequent manual finetuning of the Auto-strategy often led to further improvements.

A Computational Approach to Validating Group I Intron Subdivisions

Since 1990, no computational study has been conducted to verify the validity of the initial assignment of Group I introns into sub-groups (Michel and Westhof, 1990). Indeed, the set of ERPIN models that we built in 2007 essentially relied on the validity of these groupings (Lang et al., 2007), although omitting a distinction between sub-groups IB1, IB2, IB3, and IB4 due to substantial overlaps in predictions. In addition, we then created a "Group I, derived" sub-group for those introns that were not identified with the regular ERPIN models. In other words, derived introns do not well fit with typical consensus sequences and occasionally lack the peripheral structural domains and interactions. Two examples are presented in **Figure 1**. ERPIN models that identify the derived introns contain only subsets of the Group I core elements.

Whereas this computational intron group identification procedure has been unmatched and widely used for mt gene model inference (see Footnote 1), it is not without potential pitfalls. Its ERPIN models could not be rigorously tested for confident resolution of conflicting hits due to a lack of intron sequences at the time. Thus, using the best E-value as an *ad hoc* criterion may occasionally lead to misidentification when model predictions overlap. Secondly, contamination of ERPIN

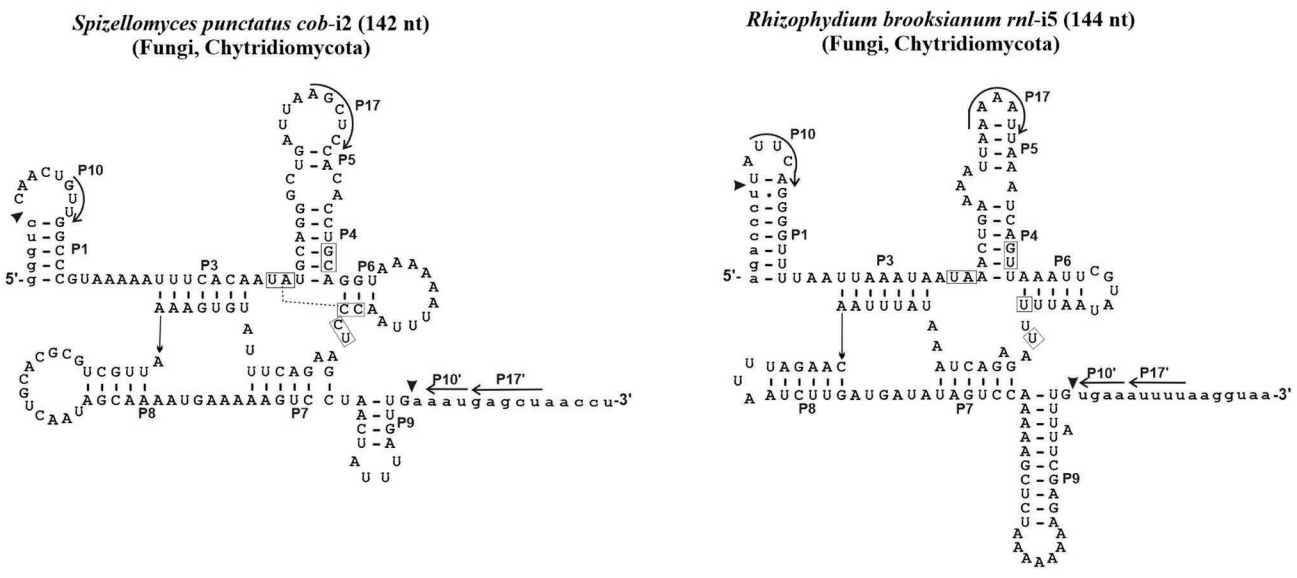


FIGURE 1 | Examples of derived Group I intron structures. With only 144 and 142 nt, the *Spizellomyces punctatus cob* intron 2 and *Rhizophydium brooksianum rnl* intron 5 do contain the universal intron core structure, but lack overall sequence conservation, are rich in A + U, and occasionally lack peripheral domains and typical interactions. Like other unusual mt introns in these two fast-evolving species, introns do not splice *in vitro* and only a small fraction (e.g., 3 out of 15 in *S. punctatus*) is identified with the ERPIN models IA1, IA3, IC1, IC2, IB, IB3, and ID as described below.

models with sequences between closely related groups (e.g., IC1 and IC2) was difficult to identify and avoid, given the small number of available sequences. Thirdly, a large fraction of group I introns contains long intron ORF insertions of more than 1,000 nt, whereas others are short and compact. The underlying intron alignments of the ERPIN models are therefore long, yet have to predict a fraction of relatively small introns. This introduces the opportunity for matches across exon boundaries. For instance, a single hit might start at the 5′ portion of a small intron and incorrectly match the 3′ of a separate, downstream intron which would effectively "bury" the bridging exon. The proportion of such potential misidentifications is estimated to be low (a few percent), but has not been rigorously quantified. It also remains to be seen in the context of proposed nested introns, or "twintrons" (Hafez and Hausner, 2015; Mukhopadhyay and Hausner, 2021), that may be more frequent than currently assumed. In fact, conflicting ERPIN predictions may be a way of pinpointing candidate twintrons.

An Updated Set of Group I Intron ERPIN Models

To address the above questions, we built a new set of ERPIN models (**Figure 2**) starting from the 87 sequences used in the originally defined Group I intron subdivision [named "Michel

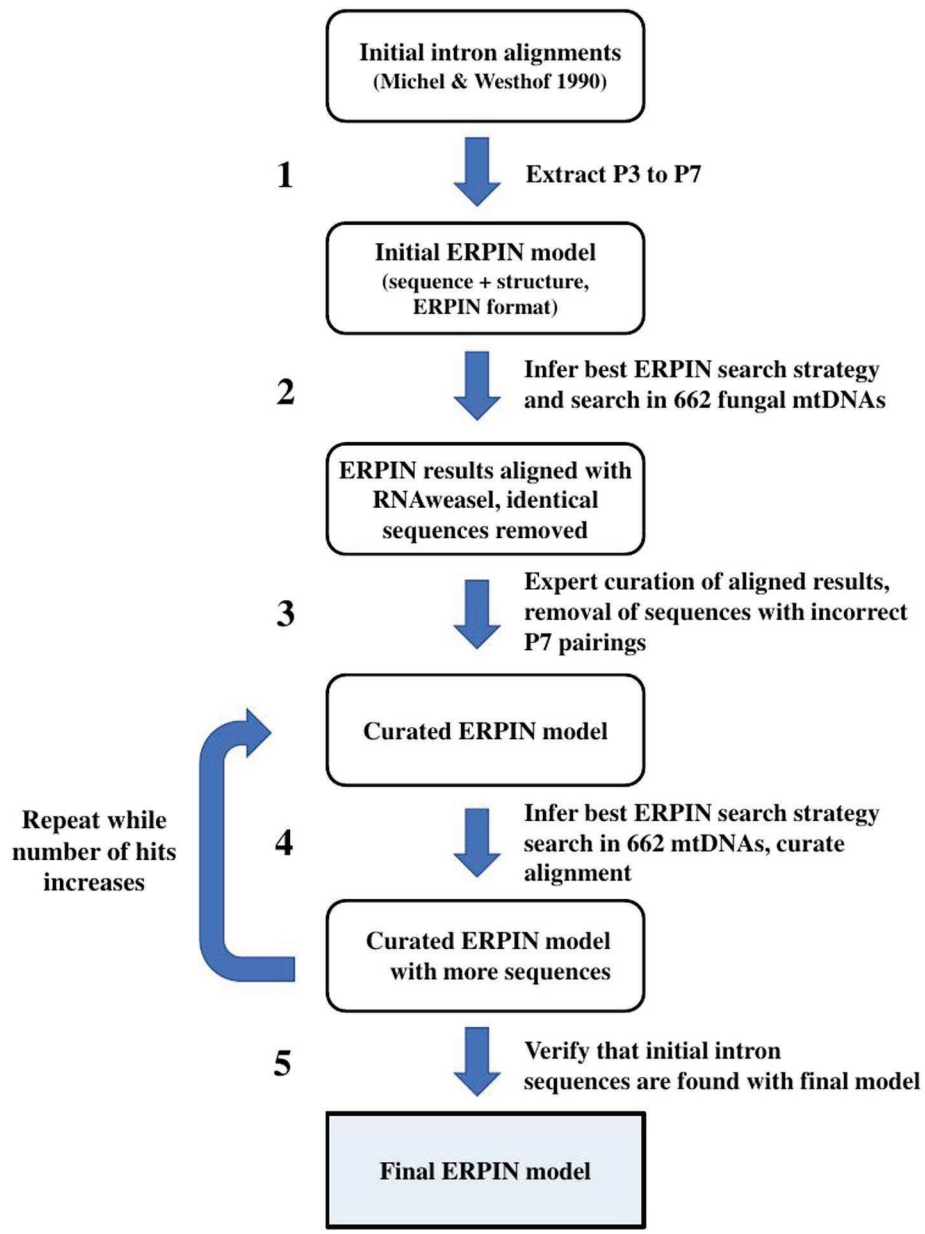


FIGURE 2 | Iterative ERPIN model construction. The procedure was applied to all separate intron sub-groups.

collection" in the following; (Michel and Westhof, 1990)]. After building structural alignments (ERPIN models) comprising P3 throughout P7 pairings (**Figure 3**), including conserved flanking sequences (if present), the set of 662 fungal mtDNA sequences (see above) was searched. Best hits were selected and added to the initial model alignment to improve model sensitivity (for details see Methods). The total numbers of predicted introns are listed in **Table 1**, together with conflicting predictions. In all instances, there was a large E-value difference between conflicting predictions, which allowed unambiguous intron group assignment.

Note that the process of building increasingly broader and detailed structural alignments depended on a program named RNAweasel,[4] which extends the functionality of ERPIN. RNAweasel aligns the resulting matches against the search model in the format of a new ERPIN model that can be directly used for subsequent searches or merged with previous alignments. In addition, it has functionalities, such as ordering of hits by E-value, removal of sequences from the alignment that are either identical, or closely related, using a phylogenetic distance measure. Finally, the RNAweasel output helps with the evaluation in providing a structured view of the search results (for an example, see **Figure 4**).

An inspection of the updated ERPIN models shows few conflicting predictions (i.e., covering the same genome loci with nearly identical coordinates), without any conflict observed for Group IA3, IC2, and ID. A small number of overlaps were observed between models IC1 and IC2 (0.6% of IC1 hits are IC2 introns), IA1 and IB (0.3% of IA1 hits are IB introns), and IB and IC1 (2.5% of IB hits are IC1 introns). Yet, in every case, a large difference in E-values allows a clear

[4]https://github.com/BFL-lab/RNAweasel

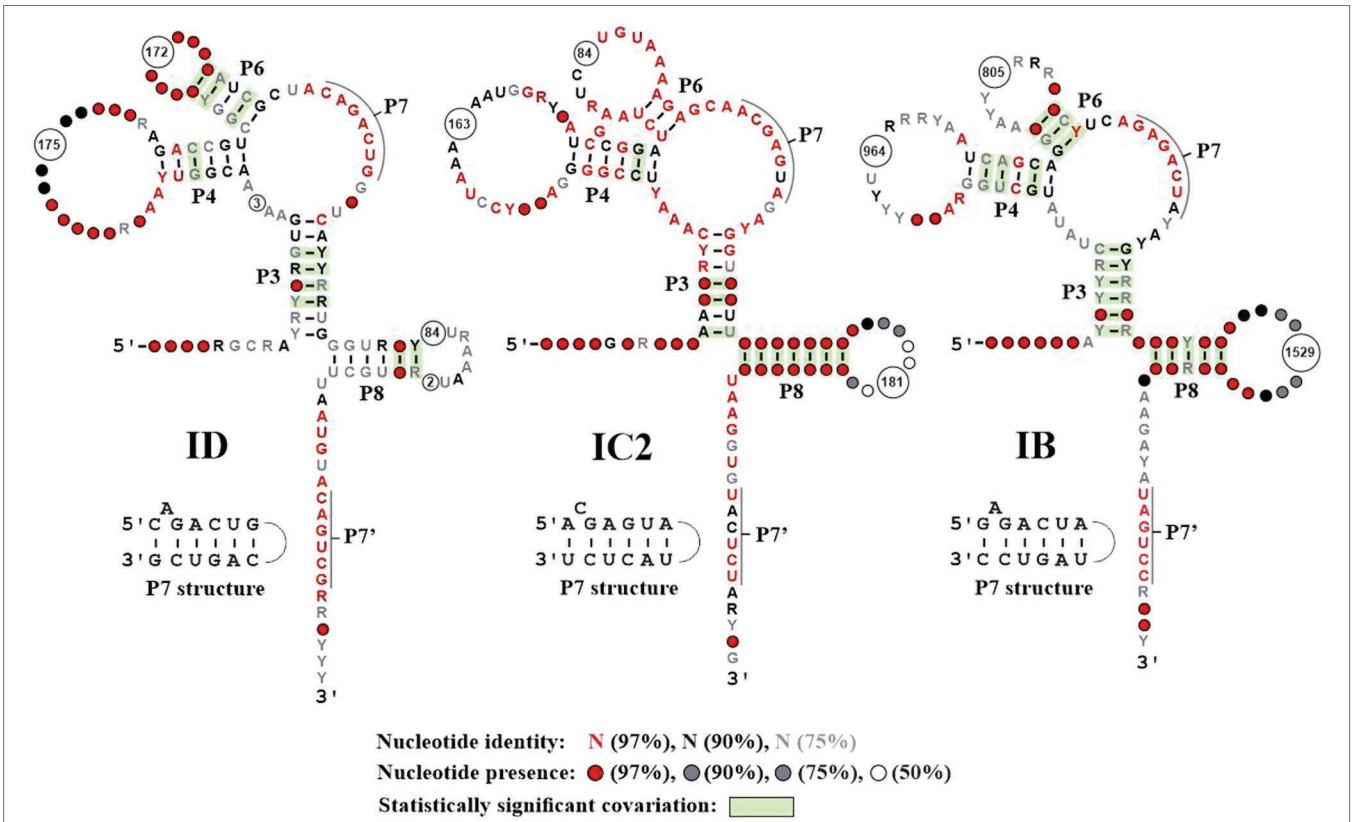


FIGURE 3 | Conserved features of ERPIN models. The three images correspond to the secondary structure drawings produced by R-scape (Rivas et al., 2020), when providing structured alignments of ERPIN models ID, IC2, and IB (the most productive in terms of numbers of hits) as input. The graphs are colored according to R-scape's conventions. As R-scape removes columns with less than 50% nucleotide presence from the graph, the number of removed columns are identified in the consensus structure as a circle with the number of removed columns inside. The P7 pseudoknot is annotated in the graph as two single-stranded regions, marked by a black line. The corresponding helical pairing is presented in a separate drawing. The number and structure of conserved elements in the remainder of ERPIN models IA1, IA3, IB3, and IC2 is essentially the same as depicted in this figure (not shown).

TABLE 1 | Number of intron predictions for distinct Group I sub-groups (in 662 mtDNAs).

	IA1	IA3	IB	IB3	IC1	IC2	ID
Total	960	78	3,582	202	337	948	1,105
E-values (e-)	(9–35)	(16–42)	(4–39)	(14–44)	(22–94)	(27–95)	(12–44)
Conflicting (true)	3(IB)	-	88(IC1)	-	2(IC2)	-	-

identification of the best-fitting intron sub-group. A special case of conflicting predictions is "**overlapping introns**" where two introns are identified in the same region but have different predicted upstream and/or downstream splicing sites (reported in **Table 2**). Again, the number of conflicts is small, occurring in only 0.7% of all total predictions. 97% of these cases are between the IB or IB3 models and other sub-groups (**Table 2**). An inspection reveals that in the majority of instances, the shorter of the conflicting alternatives is the one that is consistent with MFannot gene models.[5] In any case, we suggest that ERPIN intron predictions or MFannot gene models should be inspected by an expert before use in publications.

In the following sections, we will describe and discuss Group I models in the order ID, then IA1, IA3, IC1, and IC2, and finally IB with its four sub-groups. Starting out with the ID models will establish a basis for an unambiguous discussion on variations in the P7 motifs conservation, without ambiguity, as the result of ID searches do not conflict with any other Group I search, except for three instances of partial intron predictions (**Table 2**).

Structural Conservation of Group ID Introns

Group ID comprises a sizeable collection of mt introns (1,105, when searching against our set of 622 mtDNAs) with a predominant P7 structure 5-C_A_GACUG --- CAGUCG-3′ (**Figure 5**, upper left; for a comparison of P7 motifs among various groups, see **Figure 5**). Yet, searches also identify sequences with various different P7 motifs (**Figure 5**) that come with E-values well within the range of canonical ID introns. Among these variants, 32 have a most unusual P7 motif containing a bulged C residue (**Figure 5**; middle, left), which happens to be the predominant P7 structure of mt IC1 introns (Michel and Westhof, 1990). These 32 variant ID introns are not found with the mt IC1 ERPIN model, suggesting that the P7 structure alone is a potentially misleading feature when used for intron group identification. To exclude the possibility that our ID ERPIN model may contain sequences from other sub-groups, thus leading to potential misidentification, only those sequences with the predominant P7 structure were used to build a new ERPIN model. However, when searched against our mtDNA collection results were essentially the same, finding all variant P7 sequences listed in **Figure 5**, and within the range of E-values as introns with the canonical P7 structure (**Figure 6**). A phylogenetic analysis with all ID intron sequences is overall not well resolved (not shown), but regroups some of the ID intron members with variant P7 structures, suggesting an evolutionary process that transitions from one to another evolutionary stable P7 conformation.

Note that the total number of introns listed by P7 variant in **Figure 5** does not add up to a total of 1,105 hits. The missing 92 sequence have either variant P7 motifs not listed in the figure, or after manual inspection, the predicted sequences fail to form a proper P7 pairing. This is due to a rare ERPIN misidentification where the absence of a well-fitting match reports a spurious alternative. Despite issues with the prediction of a P7 motif in these few instances, these hits suggest the presence of true introns and are thus of some value for the inference of gene models. There are several potential reasons for this type of error, nonetheless. For instance, sequence or genome assembly error may be responsible, or presence of intron structures that exceed the

[5]https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl

```
************************************************
Condensed alignment for Bder1.epn-548.fullalignment
************************************************
There are 548 matches

Species name              E-value   # of nt     Start..Stop     Str  #  Structure
--------------------  ---------- --------- -------------------- -- --- ------------------
                                                                [[[[[[                ]]]]]] [[[                ]]]     [   [[[[[                          ]]]]] ]
Podo.comata.mt        |   2.07e-10      1241        61..265       FW   1 TGCTGA aac..(102)...caa   TCAGCA GGA aac...(36)...gga   TCC TCA G A GACTA tac...(27)...ggt   AAGATA TAGTC C
Usti.maydis.mt        |   5.18e-10      1548       130..309       RC   1 TGCTGG aaa...(70)...taa   TCAGCA GGA aac...(32)...gga   TCC TCA G A GACTA tgt...(38)...ttg   AAGATA TAGTC C
Usti.maydis.mt        |   5.18e-10      1345       152..1167      RC   1 TGCTGG aac...(57)...cga   TCAGCA GGA aat...(24)...agc   TCC TCA G A GACTA cac..(895)...atg   AAGATA TAGTC C
Mono.sp_JEL15.mt      |   8.10e-10      1488       153..343       FW   1 TGCTGG aac...(79)...caa   TCAGCA GGA aac...(30)...gga   TCC TCA G A GACTA tac...(42)...ttg   AAGATA TAGTC C
Aspe.japonicus.mt     |   8.10e-10      1560      1261..1434      FW   1 TGCTGG aac...(65)...caa   TCAGCA GGA aac...(33)...gga   TCC TCA G A GACTA aac...(36)...ctg   AAGATA TAGTC C
Mort.verticillata.mt  |   8.10e-10       484       153..325       FW   1 TGCTGG aaa...(65)...caa   TCAGCA GGA aac...(25)...gga   TCC TCA G A GACTA cac...(43)...atg   AAGATA TAGTC C
Mort.verticillata.mt  |   8.10e-10      1574      1196..1444      FW   1 TGCTGG aac...(87)...caa   TCAGCA GGA aac...(53)...gga   TCC TCA G A GACTA cac...(69)...gta   AAGATA TAGTC C
Sacc.cerevisiae.mt    |   8.10e-10      1210       158..1074      FW   1 TGCTGG aaa...(83)...taa   TCAGCA GGA taa...(46)...aga   TCC TCA G A GACTA caa..(748)...ata   AAGATA TAGTC C
Amoe.parasiticum.mt   |   8.10e-10       443       148..320       FW   1 TGCTGG gaa...(79)...caa   TCAGCA GGA aac...(30)...gga   TCC TCA G A GACTA tac...(24)...atg   AAGATA TAGTC C
Gibb.zeae.mt          |   8.10e-10      1653      1247..1521      FW   1 TGCTGG aaa...(63)...cta   TCAGCA GGA aac..(124)...gga   TCC TCA G A GACTA cgt...(48)...att   AAGATA TAGTC C
Podo.anserina.mt      |   8.10e-10      1656      1337..1530      FW   1 TGCTGG aac...(57)...caa   TCAGCA GGA aac...(66)...gga   TCC TCA G A GACTA tac...(31)...ctg   AAGATA TAGTC C
```

FIGURE 4 | RNAweasel output of ERPIN search results. Example of a formatted output from a "derived ERPIN model," demonstrating sorting of results by E-value, marking of conserved motifs in a structure definition line, marking of conserved, searched sequence blocks in upper case, and indication of the length of variable regions.

TABLE 2 | Conflicting partial intron predictions.

	IA1	IA3	IB	IB3	IC1	IC2	ID
IA1	0						
IA3	1	0					
IB	28	14	4				
IB3	0	19	6	0			
IC1	1	0	14	0	0		
IC2	0	0	13	0	0	0	
ID	1	0	2	0	0	0	0

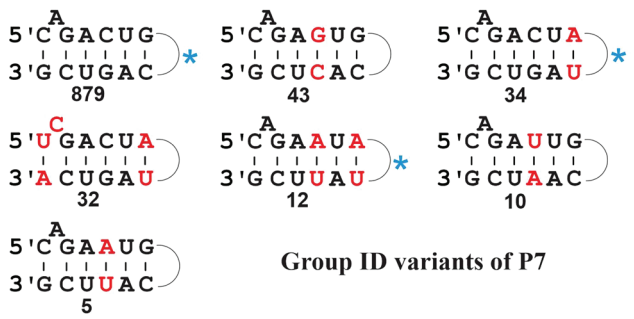


FIGURE 5 | Frequency of P7 structure variants in fungal mt Group ID introns. The numbers are compiled from the search result with the updated ID ERPIN model (see main text), identifying a total of 1,107 hits (span of E-values from 7.17e-44 to 4.94e-11); none of the respective intron sequences was identified with other group-specific searches, and all are therefore considered *bona fide* ID introns. Variant motifs with at least five identical P7 matches (numbers indicated below the structures) are listed, and nucleotides that differ from the most prevalent motif are marked in red. The most prevalent P7 motif (879 hits) is 5'-C_A_GACUG…CAGUCG-3'. The blue asterisk beside motifs indicates listing of the same motif in the Michel and Westhof compilation (Michel and Westhof, 1990).

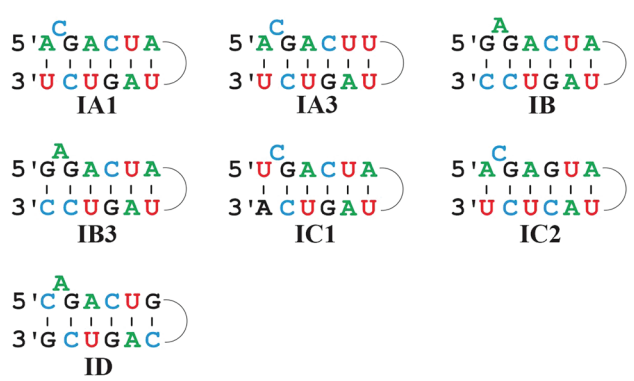


FIGURE 6 | Comparison of predominant P7 motifs in mt Group I intron sub-groups. Most frequent P7 motifs in mt Group I intron sub-groups. Note that IB and IB3 have the same P7 motif, but that they are clearly distinct otherwise and identified with separate ERPIN models. IA3 introns have a frequent alternative motif A_C_GACUG … UAGUC_U (i.e., with a single nucleotide difference to the one depicted above).

length of the respective ERPIN models, or genetic rearrangements due to intron mobility that may introduce sequence duplication and recombination with other intron sub-groups. At about 0.7% of the total count, this constitutes a tolerable degree of uncertainty. Yet, it reminds us that intron identification is but another tool to provide evidence for the detection and resolution of inconsistencies, as part of a more complete gene modeling procedure.

Distribution of Group IA1, IA3, IC1, and IC2 Introns

In contrast to sub-group IA1, IC1, and IC2 introns that are frequently identified in mtDNAs (**Table 1**), sub-group IA3 introns are remarkably rare. In fact, the Michel collection only contains four plastid IA3 introns and no mt representative. It was therefore surprising to find as many as 78 strong mt hits (in the E-value range between 4.71e-16 and 1.81e-42) distributed across Fungi, but with a strong preference for basidiomycetes. Notably, 18 of the 78 sequences do not have the predominant bulged C in the P7 motif but instead an A, and these sequences closely regroup in the less specific e-16 range of E-values. When separating sequences in the IA3 ERPIN model by these two P7 motifs to form two distinct sub-models, these identify different but somewhat overlapping sets of introns. If this result may be taken to suggest a separate, new (IA4?) sub-group remains to be clarified, once more mtDNAs become available.

Despite some structural similarities, sub-group IC1 and IC2 ERPIN models are clearly distinct in terms of mt intron identification, with only marginal conflicts. The few conflicting predictions are separated by substantial E-value differences, thus allow for unambiguous sub-group assignment. In this context, it is interesting to test entries in the GISSD database, which lists a total of 1789 sequences, with close to half (837; most in nuclear eukaryotic rDNAs) labeled as IC1. Curiously, Group IC1 entries in the database do not list any of the mt IC1 introns in the Michel collection that served as our starting point for developing an mt IC1 ERPIN model. When searching the Michel collection with the IC1 model it identifies all (and only) mt IC1 entries. It was therefore interesting to test if the mt ERPIN model would also identify introns in nuclear rDNAs (at such a large evolutionary distance), and that despite a strong difference in nucleotide bias (mt sequences most A + T rich, vs. nuclear sequences G + C rich). The results show clear identification of 479 out of the 837 listed IC1 introns, with E-values ranging from 5.40e-25 to 6.53e-49, and without modification of the ERPIN search parameters. The identification of the remainder of nuclear IC1 introns was possible only after transitioning to a nuclear-sequence specific ERPIN model, however. Evidently, these two lineages of introns (mt vs. nuclear) have undergone a separate evolutionary path, under different genetic constraints.

Confirmation of Group IB3, yet No Computationally Distinct IB1, IB2, and IB4 Sub-Groups

In an attempt to follow up on the proposal of four separate IB sub-groups, ERPIN models were developed according to our protocol and tested for potentially conflicting predictions. Whereas IB3 turns out to be a small, clearly separate group, the other three sub-groups overlap substantially in predictions (**Figure 7**), without a possibility for separation based on E-values. The IB1, IB2, and IB4 models properly identified representatives in the Michel collection with high scores, confirming the validity of the model building procedure. Yet, searches against the GISSD intron collection resulted in matches with barely recognizable distinction between the sub-groups. We conclude that IB1, IB2, and IB4 are too closely related for establishing distinct sub-groups, and have therefore joined them into a IB super-group as reported in Tables and Figures above.

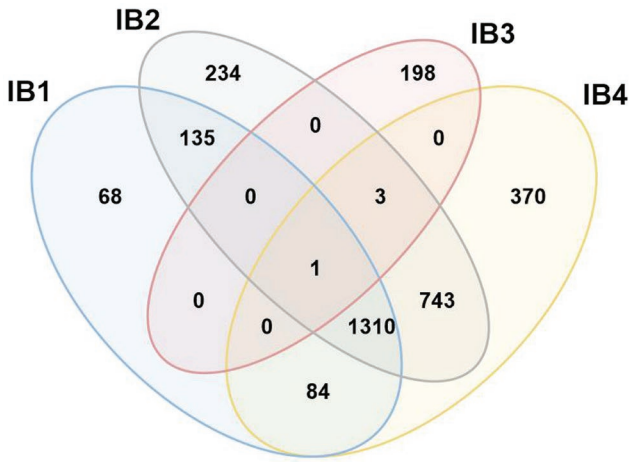


FIGURE 7 | Venn diagram demonstrating conflicting hits among IB sub-groups. Search results with ERPIN models specific for the four Group IB sub-groups [as proposed in (Michel and Westhof, 1990)] were analyzed for conflicting hits, with respective numbers displayed in a Venn diagram. With the notable exception of IB3 which has only 4 conflicts out of a total of 202 intron sequences, IB1,2 and 4 overlap substantially in identification, without a possibility to distinguish by E-value. We therefore combine IB1,2 and 4 into a formal group IB that will co-exist with a separate IB3.

Cm Analyses Based on ERPIN Structural Alignments

A final point of interest is the conversion of ERPIN models into covariance models, for comparing the performance and precision of the two conceptually very different approaches. From a technical point of view, this conversion is easy as the alignment in ERPIN models is in fasta format, which is readily reformatted into Stockholm format (.sto) required for cmbuild [infernal; (Nawrocki et al., 2009)]. Because the ERPIN intron models contain pseudoknots that are encoded by a numbering scheme, the respective structure line for the sto format was translated in WUSS format. The current Infernal version does not use pseudoknot information and treats pseudoknot pairings as conserved at the nucleotide level only. For testing purposes, we chose two ERPIN models, ID and IA3 that differ substantially in length and in the degree of relatedness among the aligned sequences. The ID alignment is relatively short (560 nt positions), compact, and moderately conserved, whereas IA3 is long (3,042 nt), with very large insertions (up to 2011 nt) and highly conserved. The outcome is disappointing in the sense that both CMs have a smaller number of complete matches (776 for ID; 45 for IA3) compared to ERPIN model searches. The total numbers of matches (better than 1.0e-2) with CMs is substantially higher than the total number of respective ERPIN matches, indicating a better potential to identify more introns, although missing the ability to properly align the complete CM to the genome sequences. To our interpretation, the CM approach would profit from reorganizing the search algorithm, from a strictly HMM-like scanning from 5′ to 3′ toward a modular motif-driven approach used by ERPIN—which may at the same time resolve the issue of introducing pseudoknot information in CM searches.

CONCLUDING REMARKS

Here we present an update of our previous work (Lang et al., 2007), and a more detailed description on the identification of mt Group introns, in light of new publicly data available, using ERPIN models. The update makes progress on some remaining issues from the previous work, extends the accuracy of the models, and sheds light on Group IB introns. Enhanced model sensitivity and specificity was achieved through two means. First, multiple sequence alignments of intron sequences were significantly extended by virtue of newly available data. Second, we developed a more systematic approach to curation of alignments, to exclude sequences that do not belong with a given sub-group (increased risk of incorrect identification). While an overall increase in model sensitivity was achieved, sub-groups IB1, IB2, and IB4 were found to be too closely related, which hampered model specificity, suggesting that IB1, IB2, and IB4 may be dismissed as sub-groups. On the other hand, Group IB3 introns were found to be sufficiently distinct to build a highly sensitive and specific ERPIN model. The current intron predictors are expected to improve the gene modeling of the MFannot tool as well as provide more precise structural intron information.

A remaining gray zone of Group I intron identification pertains to those that appear less well-structured, or "derived." Our attempts to establish clearly distinct ERPIN models that would include those derived introns have been so far without success. Likewise, our attempt to transition from ERPIN to expectedly more sensitive CM searches has come with mixed success. It has provided more hits than ERPIN, however at the cost of a reduced number of full intron hits and a large portion of partial hits with borderline scores. To our assessment, a modification of the CM approach that allows modular search of conserved motifs or regions might be a potential solution, which would at the same time allow for the use of pseudoknot information.

Evidently, a continued search for additional distinct ERPIN sub-groups would be in order, but its algorithm is of little help with developing new structured alignments as required. For this, a more modular Infernal version in combination with primary sequence alignment [e.g., Muscle (Edgar, 2004) in combination with HMM searches (Eddy, 2011)] with secondary structure modeling [e.g., RNAalifold (Bernhart et al., 2008) or R-scape (Rivas et al., 2017)] would be preferable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SP investigated conflicting intron predictions. CM was in charge of covariance analyses and the design of R-scape graphics. FF-B was involved in developing ERPIN search strategies. PR designed and coded RNAweasel. MS provided informatics and coding support. BL established optimized ERPIN models. All co-authors participated in discussing and writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors acknowledge generous support by the Natural Sciences and Engineering Research Council of Canada (NSERC grant numbers RGPIN-2014-05286 and RGPIN-2017-05411; plus an NSERC internship scholarship for CM) and by the Fond de Recherche Nature et Technologie, Quebec.

ACKNOWLEDGMENTS

The authors acknowledge support in 2D modeling and design of mt intron structures by M-J. Laforest and suggestions for building improved CMs with large variable inserts by Eric Nawrocki.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.866187/full#supplementary-material

REFERENCES

Augustin, S., Muller, M. W., and Schweyen, R. J. (1990). Reverse self-splicing of group II intron RNAs in vitro. *Nature* 343, 383–386. doi: 10.1038/343383a0

Bassi, G. S., de Oliveira, D. M., White, M. F., and Weeks, K. M. (2002). Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc. Natl. Acad. Sci. U. S. A.* 99, 128–133. doi: 10.1073/pnas.012579299

Belfort, M., and Bonocora, R. P. (2014). Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods Mol. Biol.* 1123, 1–26. doi: 10.1007/978-1-62703-968-0_1

Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., and Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinform.* 9:474. doi: 10.1186/1471-2105-9-474

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes - anything goes. *Trends Genet.* 19, 709–716. doi: 10.1016/j.tig.2003.10.012

Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E., and Slonimski, P. P. (1983). An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell* 35, 733–742. doi: 10.1016/0092-8674(83)90106-X

Cech, T. R., Tanner, N. K., Tinoco, I. Jr., Weir, B. R., Zuker, M., and Perlman, P. S. (1983). Secondary structure of the *Tetrahymena* ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. *Proc. Natl. Acad. Sci. U. S. A.* 80, 3903–3907. doi: 10.1073/pnas.80.13.3903

Cho, Y., and Palmer, J. D. (1999). Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the *Araceae* family. *Mol. Biol. Evol.* 16, 1155–1165. doi: 10.1093/oxfordjournals.molbev.a026206

Cinget, B., and Bélanger, R. R. (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, 1252–1260. doi: 10.1080/15476286.2020.1763024

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Cummings, D. J., McNally, K. L., Domenico, J. M., and Matsuura, E. T. (1990a). The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Curr. Genet.* 17, 375–402. doi: 10.1007/BF00334517

Cummings, D. J., Michel, F., Domenico, J. M., and McNally, K. L. (1990b). Mitochondrial DNA sequence analysis of the cytochrome oxidase subunit II gene from *Podospora anserina*. A group IA intron with a putative alternative splice site. *J. Mol. Biol.* 212, 287–294. doi: 10.1016/0022-2836(90)90125-6

Dujon, B. (1980). Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell* 20, 185–197. doi: 10.1016/0092-8674(80)90246-9

Dujon, B. (1989). Group I introns as mobile genetic elements: facts and mechanistic speculations--a review. *Gene* 82, 91–114. doi: 10.1016/0378-1119(89)90034-6

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Ferat, J. L., and Michel, F. (1993). Group II self-splicing introns in bacteria. *Nature* 364, 358–361. doi: 10.1038/364358a0

Gautheret, D., and Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* 313, 1003–1011. doi: 10.1006/jmbi.2001.5102

Gray, M. W., Burger, G., and Lang, B. F. (1999). Mitochondrial evolution. *Science* 283, 1476–1481. doi: 10.1126/science.283.5407.1476

Hafez, M., and Hausner, G. (2015). Convergent evolution of twintron-like configurations: one is never enough. *RNA Biol.* 12, 1275–1288. doi: 10.1080/15476286.2015.1103427

Hausner, G., Hafez, M., and Edgell, D. R. (2014). Bacterial group I introns: mobile RNA catalysts. *Mob. DNA* 5:8. doi: 10.1186/1759-8753-5-8

Henke, R. M., Butow, R. A., and Perlman, P. S. (1995). Maturase and endonuclease functions depend on separate conserved domains of the bifunctional protein encoded by the group I intron aI4 alpha of yeast mitochondrial DNA. *EMBO J.* 14, 5094–5099. doi: 10.1002/j.1460-2075.1995.tb00191.x

Huang, H. R., Rowe, C. E., Mohr, S., Jiang, Y., Lambowitz, A. M., and Perlman, P. S. (2005). The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* 102, 163–168. doi: 10.1073/pnas.0407896101

Kreike, J., Schulze, M., Ahne, F., and Lang, B. F. (1987). A yeast nuclear gene, MRS1, involved in mitochondrial RNA splicing: nucleotide sequence and mutational analysis of two overlapping open reading frames on opposite strands. *EMBO J.* 6, 2123–2129. doi: 10.1002/j.1460-2075.1987.tb02479.x

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31, 147–157. doi: 10.1016/0092-8674(82)90414-7

Lambert, A., Fontaine, J. F., Legendre, M., Leclerc, F., Permal, E., Major, F., et al. (2004). The ERPIN server: an interface to profile-based RNA motif identification. *Nucleic Acids Res.* 32, W160–W165. doi: 10.1093/nar/gkh418

Lambowitz, A. M., and Zimmerly, S. (2004). Mobile group II introns. *Annu. Rev. Genet.* 38, 1–35. doi: 10.1146/annurev.genet.38.072902.091600

Lang, B. F. (2013). *Mitochondrial Genomes in Fungi*. New York: Springer.

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397. doi: 10.1146/annurev.genet.33.1.351

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

Liu, W., Cai, Y., Zhang, Q., Chen, L., Shu, F., Ma, X., et al. (2020). The mitochondrial genome of *Morchella importuna* (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames and repetitive sequences. *Int. J. Biol. Macromol.* 143, 373–381. doi: 10.1016/j.ijbiomac.2019.12.056

Michel, F., and Dujon, B. (1983). Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J.* 2, 33–38. doi: 10.1002/j.1460-2075.1983.tb01376.x

Michel, F., and Ferat, J. L. (1995). Structure and activities of group II introns. *Annu. Rev. Biochem.* 64, 435–461. doi: 10.1146/annurev.bi.64.070195.002251

Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989). The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* 342, 391–395. doi: 10.1038/342391a0

Michel, F., Jacquier, A., and Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64, 867–881. doi: 10.1016/S0300-9084(82)80349-0

Michel, F., and Lang, B. F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature* 316, 641–643. doi: 10.1038/316641a0

Michel, F., and Westhof, E. (1990). Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216, 585–610. doi: 10.1016/0022-2836(90)90386-Z

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

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Nawrocki, E. P., Jones, T. A., and Eddy, S. R. (2018). Group I introns are widespread in archaea. *Nucleic Acids Res.* 46, 7970–7976. doi: 10.1093/nar/gky414

Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25, 1335–1337. doi: 10.1093/bioinformatics/btp157

Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., et al. (1993). Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. *Nucleic Acids Res.* 21, 1297–1305. doi: 10.1093/nar/21.5.1297

Osigus, H. J., Eitel, M., and Schierwater, B. (2017). Deep RNA sequencing reveals the smallest known mitochondrial micro exon in animals: The placozoan *cox1* single base pair exon. *PLoS One* 12:e0177959. doi: 10.1371/journal.pone.0177959

Paquin, B., Laforest, M. J., Forget, L., Roewer, I., Wang, Z., Longcore, J., et al. (1997). The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr. Genet.* 31, 380–395. doi: 10.1007/s002940050220

Qiu, Y. L., Cho, Y., Cox, J. C., and Palmer, J. D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394, 671–674. doi: 10.1038/29286

Rivas, E., Clements, J., and Eddy, S. R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* 14, 45–48. doi: 10.1038/nmeth.4066

Rivas, E., Clements, J., and Eddy, S. R. (2020). Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics* 36, 3072–3076. doi: 10.1093/bioinformatics/btaa080

Schäfer, B., Gan, L., and Perlman, P. S. (2003). Reverse transcriptase and reverse splicing activities encoded by the mobile group II intron cobI1 of fission yeast mitochondrial DNA. *J. Mol. Biol.* 329, 191–206. doi: 10.1016/S0022-2836(03)00441-8

Schäfer, B., Merlos-Lange, A. M., Anderl, C., Welser, F., Zimmer, M., and Wolf, K. (1991). The mitochondrial genome of fission yeast: inability of all introns to splice autocatalytically, and construction and characterization of an intronless genome. *Mol. Gen. Genet.* 225, 158–167. doi: 10.1007/BF00282654

Shub, D. A., Gott, J. M., Xu, M. Q., Lang, B. F., Michel, F., Tomaschewski, J., et al. (1988). Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 85, 1151–1155. doi: 10.1073/pnas.85.4.1151

Turmel, M., Mercier, J. P., and Cote, M. J. (1993). Group I introns interrupt the chloroplast *psaB* and *psbC* and the mitochondrial *rrnL* gene in *Chlamydomonas*. *Nucleic Acids Res.* 21, 5242–5250. doi: 10.1093/nar/21.22.5242

Waring, R. B., Davies, R. W., Scazzocchio, C., and Brown, T. A. (1982). Internal structure of a mitochondrial intron of *Aspergillus nidulans*. *Proc. Natl. Acad. Sci. U. S. A.* 79, 6332–6336. doi: 10.1073/pnas.79.20.6332

Wheeler, T. J., Clements, J., and Finn, R. D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinform.* 15:7. doi: 10.1186/1471-2105-15-7

Zhou, Y., Lu, C., Wu, Q. J., Wang, Y., Sun, Z. T., Deng, J. C., et al. (2008). GISSD: group I intron sequence and structure database. *Nucleic Acids Res.* 36(suppl_1), D31–D37. doi: 10.1093/nar/gkm766

Zimmerly, S., and Semper, C. (2015). Evolution of group II introns. *Mob. DNA* 6:7. doi: 10.1186/s13100-015-0037-5

Zubaer, A., Wai, A., and Hausner, G. (2018). The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. *Sci. Rep.* 8:17591. doi: 10.1038/s41598-018-35926-y

Conflict of Interest: The 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fmicb.2022.821638

Mitochondrial Transcription of Entomopathogenic Fungi Reveals Evolutionary Aspects of Mitogenomes

*Stylianos P. Varassas and Vassili N. Kouvelis**

Department of Genetics and Biotechnology, Faculty of Biology, National and Kapodistrian University of Athens, Athens, Greece

Entomopathogenic fungi and more specifically genera *Beauveria* and *Metarhizium* have been exploited for the biological control of pests. Genome analyses are important to understand better their mode of action and thus, improve their efficacy against their hosts. Until now, the sequences of their mitochondrial genomes were studied, but not at the level of transcription. Except of yeasts and *Neurospora crassa*, whose mt gene transcription is well described, in all other Ascomycota, i.e., Pezizomycotina, related information is extremely scarce. In this work, mt transcription and key enzymes of this function were studied. RT-PCR experiments and Northern hybridizations reveal the transcriptional map of the mt genomes of *B. bassiana* and *M. brunneum* species. The mt genes are transcribed in six main transcripts and undergo post-transcriptional modifications to create single gene transcripts. Promoters were determined in both mt genomes with a comparative *in silico* analysis, including all known information from other fungal mt genomes. The promoter consensus sequence is 5′-ATAGTTATTAT-3′ which is in accordance with the definition of the polycistronic transcripts determined with the experiments described above. Moreover, 5′-RACE experiments in the case of premature polycistronic transcript *nad*1-*nad*4-*atp*8-*atp*6 revealed the 5′ end of the RNA transcript immediately after the *in silico* determined promoter, as also found in other fungal species. Since several conserved elements were retrieved from these analyses compared to the already known data from yeasts and *N. crassa*, the phylogenetic analyses of mt RNA polymerase (Rpo41) and its transcriptional factor (Mtf1) were performed in order to define their evolution. As expected, it was found that fungal Rpo41 originate from the respective polymerase of T7/T3 phages, while the ancestor of Mtf1 is of alpha-proteobacterial origin. Therefore, this study presents insights about the fidelity of the mt single-subunit phage-like RNA polymerase during transcription, since the correct identification of mt promoters from Rpo41 requires an ortholog to bacterial sigma factor, i.e., Mtf1. Thus, a previously proposed hypothesis of a phage infected alpha-proteobacterium as the endosymbiotic progenitor of mitochondrion is confirmed in this study and further upgraded by the co-evolution of the bacterial (Mtf1) and viral (Rpo41) originated components in one functional unit.

Keywords: mitochondrial (mt) transcription, entomopathogenic fungi, mt RNA polymerase, RPO41, mt transcription factor, MTF1, mt promoters

OPEN ACCESS

Edited by:
William Martin, University of Dusseldorf Medical School, Germany

Reviewed by:
Alexander N. Ignatov, Peoples' Friendship University of Russia, Russia
Arthur Gruber, University of São Paulo, Brazil

****Correspondence:***
Vassili N. Kouvelis
kouvelis@biol.uoa.gr

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *24 November 2021*
Accepted: *22 February 2022*
Published: *21 March 2022*

Citation:
Varassas SP and Kouvelis VN (2022) Mitochondrial Transcription of Entomopathogenic Fungi Reveals Evolutionary Aspects of Mitogenomes. Front. Microbiol. 13:821638. doi: 10.3389/fmicb.2022.821638

INTRODUCTION

Mitochondria are the semiautonomous powerhouses of the cell that carry their own small genome that does not encode for all the products needed by mitochondria (Burger et al., 2003). Especially in fungi, mitogenomes show such a great diversity regarding genome size, intergenic regions and intron content (Mukhopadhyay and Hausner, 2021; Yildiz and Ozkilinc, 2021) that can be exploited for species and strain typing (Kortsinoglou et al., 2019, 2020) and other phylogenetic and evolutionary studies (Hausner, 2003; Aguileta et al., 2014; Fonseca et al., 2020; Glare et al., 2020; Megarioti and Kouvelis, 2020). However, almost within all fungal mt genomes, genes encoding rRNA (*rns* and *rnl*), tRNAs (*trn*) and proteins involved in mitochondrial ribosome assembly (*rps*3) and oxidative phosphorylation [i.e., subunits of ATP synthase (atp6, 8 and 9), apocytochrome b (*cob*) cytochrome c (*cox*1-3) and NADH dehydrogenase (*nad*1-6 and *nad*4L)] are commonly found (Paquin et al., 1997; Kouvelis et al., 2004). In only two cases this gene content is extensively different. Specifically, in several fungal species *rps*3 is missing or located in the nuclear genome (Korovesi et al., 2018) and *nad* genes cannot be found in the mitogenomes of the family Saccharomycetaceae of Saccharomycotina (Christinaki et al., 2021).

Therefore, it is expected that all functions of the mitochondria, like the transcription of their mt genes, to remain conserved. Until now, knowledge for this function in fungal mt genomes relies mostly on studies performed in yeasts and fungal model organisms like *Neurospora crassa* and *Schizosaccharomyces pombe* (Burger et al., 1985; Kleidon et al., 2003; Schäfer, 2005; Schäfer et al., 2005). In detail, mt transcription requires promoters which are participating in the simultaneous transcription of several genes in one transcript (polycistronic transcripts), as a result of the endosymbiotic origin of the mitochondria from a-proteobacteria, in which their genome is organized in operons under the regulation of one promoter (Martin et al., 2015). Based on yeasts studies, mt promoters contain an "A-T enriched" consensus sequence of 9 bp, which is recognized by the mitochondrial transcription machinery (Christianson and Rabinowitz, 1983; Schäfer et al., 2005; Deshpande and Patel, 2014). More specifically, in yeasts an mt promoter is characterized by the 5′-(−8)ATATAAGTA(+ 1)-3′ sequence, where + 1 represents the transcription start site (Biswas et al., 1985; Schinkel et al., 1986; Biswas, 1999; Kim et al., 2012). In *Neurospora crassa*, the mitochondrial promoter has a modified AT-rich consensus sequence, i.e., 5′- TTAG(A/T)RR(G/T)(G/C)N(A/T)-3′ (Kubelik et al., 1990; Kleidon et al., 2003). Mt promoters are dispersed throughout the mt genome. For example, in *Saccharomyces cerevisiae*, the mitochondrial transcription initiates in more than 10 different areas, each having this consensus as a promoter sequence (Costanzo and Fox, 1990; Pel and Grivell, 1993; Grivell, 1995; Gagliardi et al., 2004). After mt transcription produced the initial polycistronic transcripts, they are further processed and converted to monocistronic (mRNA, tRNA, rRNA) by being digested at the 5′ and 3′ untranslated regions (UTRs). This transcript maturation has been proven experimentally not only in *S. cerevisiae*, but additionally, in *Neurospora crassa* and *Schizosaccharomyces pombe* (Kennell and Lambowitz, 1989; Kubelik et al., 1990; Dieckmann and Staples, 1994; Schäfer, 2005). Whenever an intron was found within the mt gene, introns were self-spliced (Cech, 1990; Michel and Ferat, 1995), leading, thus, the whole transcription procedure to the final maturation of the mt RNAs, i.e., "single intronless gene transcript." The mechanism of transcription termination in fungal mt genomes is not known (Lipinski et al., 2010), and only in vertebrates (Clayton, 1991), there are specific transcription termination sequences to enable the release of the transcript from the mtDNA. Alternatively, the formation of stem-loop structure in the native mitochondrial RNA and/or tRNAs might play the signal of transcription's termination (Breitenberger et al., 1985; Christianson and Clayton, 1988; Clayton, 1991; Schäfer et al., 2005). The mature transcripts bear no cap nor polyadenylation (Rorbach et al., 2014).

Regarding the most important proteins implicated in fungal mitochondrial transcription, two proteins, i.e., mtRNA polymerase (mtRNAP or Rpo41) and Mtf1 play the most significant roles (Lipinski et al., 2010; De Wijngaert et al., 2021). Both proteins are encoded by nuclear genes, i.e., *rpo*41 and *mtf* 1, as the former is the catalytic component of the transcription and the latter is the sole needed transcription factor for the orderly function of the polymerase, in contrast to the nuclear polymerases for which many different transcription factors are needed (Yang et al., 2015). It is known that RNA polymerase (mtRNAP) is homologous with the respective polymerase of bacteriophage T3/T7 (Masters et al., 1987; McAllister and Raskin, 1993; Filée and Forterre, 2005). Unlike the RNAP of the T7 phage, which is a single subunit RNA polymerase (ssuRNAP) that alone catalyzes all stages of transcription, mtRNAP is dependent from transcription factors for the initiation of this process (Bonawitz et al., 2006; Jiang et al., 2011). Similarly, archaeal, and fungal nuclear RNA polymerases comprise many modules (multi-subunit RNAP; msuRNAP) and depend on a large number of transcription factors (Borukhov and Severinov, 2002). Rpo41 in yeasts consists of a single enzyme subunit, divided into several domains, including the C-terminal domain (CTD), the N-terminal domain (NTD), and the N-terminal extension (NTE) (Rodeheffer et al., 2001; Deshpande and Patel, 2012). The Mtf1 is a compact protein of 43 kDa, that has an N-terminal region, C-terminal region and a flexible C-terminal tail (Schubot et al., 2001; Deshpande and Patel, 2012). The C-terminal tail is very important as it interacts with the promoter during the creation of the transcription initiation complex (Savkina et al., 2010). The N-terminal domain contains the S-adenosyl-methionine (SAM)—binding site, which is usually found in methyltransferases, although the SAM-binding site is not essential for transcription (Cotney and Shadel, 2006). Nevertheless, the mt transcription pre-initiation complex of Rpo41 and Mtf1 plays an active role in the melting of the DNA and the local denaturation of the promoter region (Sologub et al., 2009; Deshpande and Patel, 2014; Ramachandran et al., 2016). The denatured nucleotides from mtRNA polymerase's activity can be found at positions −4 to + 2 in the promoter of the mitochondrial gene (Paratkar and Patel, 2010; Velazquez et al., 2015; Basu et al., 2020).

While the majority of research on mt genome diversity and its functions has been focused on yeasts (e.g., Freel et al., 2015; Kolondra et al., 2015), it is important to expand further the studies to other subphyla of Ascomycota, specifically in the Pezizomycotina (Pantou et al., 2008) in order to fully clarify the similarities of the mt gene expression mechanisms. Entomopathogenic fungi like the Hypocrealean species *Metarhizium brunneum* and *Beauveria bassiana* may act as model organisms for studying the transcription mechanism of mt genes for several reasons. Firstly, their mt genomes are already known and analyzed (Ghikas et al., 2010; Kortsinoglou et al., 2020) in addition to their whole genomes (Gao et al., 2011; Saud et al., 2021). Secondly, both species have been used as Biological Control Agents (BCAs), and thus, they have been proposed as safe alternatives for the protection of several different crops worldwide (Butt et al., 2001; Typas and Kouvelis, 2012). Gaining insights into the functional mechanisms of their mt genomes may provide a starting point for future genetic modifications of these genomes, with a final aim of improving the efficacy of their entomopathogenic activity against the pests which destroy the crops.

The scientific and economical interest for exploiting these two entomopathogenic fungi (Gao et al., 2011; Kouvelis et al., 2011; Xiao et al., 2012), the need for better knowledge of BCAs (Clarkson and Charnley, 1996; Strasser et al., 2000; Vey et al., 2001; Shahid et al., 2012; Butt et al., 2016; Lovett and Leger, 2016) and the lack of genetic data at the mitochondrial genome processes of the above entomopathogenic fungi (Kouvelis et al., 2004; Ghikas et al., 2006, 2010) were the motivation for studying the mechanisms and prime enzymes of fungal mitochondrial transcription (mtRNA processing). Therefore, in this study the promoters of the mt genomes for these two Hypocrealean species and their polycistronic transcripts will be determined. In addition, the phylogenetic evolution of the two nuclear encoded proteins that are crucial for the mt transcription, like the mtRNA polymerase (Rpo41) and its transcription factor Mtf1, will provide important knowledge through the comparative analysis with the already known data from yeasts. Thus, the evolution of a mitochondrial process like transcription will be examined from both the "structural" (promoters, transcripts, maturation of polycistronic transcripts) and the "functional" (key protein enzymes involved, like Rpo41 and Mtf1) aspects in order to create a hypothesis which explains the current form of mt transcription.

MATERIALS AND METHODS

Strains Used, Growth Conditions and DNA/RNA Extraction

Metarhizium brunneum strain ARSEF 3297 (from USDA-ARS Collection of Entomopathogenic Fungal Culture, Ithaca, NY, United States) and *Beauveria bassiana* strain ATHUM 4946 (from ATHUM culture collection of fungi, Athens, Greece) were used for determining the primary polycistronic transcripts of their mitochondrial genomes. Fungi were grown on a rich complete Potato Dextrose Agar (PDA) liquid medium, with the addition of 2 gl^{-1} casein hydrolyzate, malt extract, yeast extract, and mycological peptone (Kouvelis et al., 2008a). For solid media, 1.5% agar was added. All strains used were derived from single conidia grown on PDA plates (11 days old). Shake flask cultures were grown at 25°C and after 4–5 days, mycelia were collected by vacuum filtration, lyophilized for 4 days and crushed in liquid nitrogen using a mortar and pestle. Lyophilized, ground mycelia were maintained at −80°C. Isolation of total genomic and mitochondrial DNA was performed as previously described (Kouvelis et al., 2008b). Total cellular RNA was isolated from 50 to 100 mg lyophilized, grounded mycelium using the TRIzol™ Reagent (Invitrogen, Waltham, MA), following the manufacturer's instructions. Total RNA was eluted in 150 μl of RNase free water (DEPC-treated), treated with DNAse I and stored at –80°C.

Mitochondrial Genome Annotation and Analysis

The complete nucleotide sequence of the mitochondrial DNA from the entomopathogenic fungus *Metarhizium brunneum* ARSEF3297 appears in the GenBank Nucleotide database under accession number NW_014574670.1 (Contig: AZNG01000047) (Hu et al., 2014). In this work, the mitochondrial genome of strain ARSEF3297 was retrieved and annotated (**Supplementary Figure 1** and **Supplementary Table 1**) as described previously (Ghikas et al., 2006). Specifically, the protein coding and the ribosomal (rRNA) genes were identified using BLASTx and BLASTn (Altschul et al., 1990), respectively, after comparisons with known related sequences and mostly with the other known mt genome of *M. brunneun* ARSEF 4556 (**Supplementary Table 2**; Kortsinoglou et al., 2020). The *trn* genes were detected using the online software tRNAScan-SE (Chan and Lowe, 2019) and intron characterization was performed using RNAweasel and MFannot (Lang et al., 2007).

In silico Prediction of Promoters

The complete mt genomes of *M. brunneum* ARSEF 3297 (the leading strain of this work), *M. brunneum* ARSEF 4556 (Kortsinoglou et al., 2020), *M. acridum* CQMa102 (Gao et al., 2011), and of *B. bassiana* Bb147 (Ghikas et al., 2010) were aligned using MegAlign program of Lasergene-DNASTAR suite by employing ClustalW (Larkin et al., 2007) with default parameters. Emphasis was given in the 3′ end of intergenic regions at the 5′ end of all mt genes in order to retrieve the conserved regions. In addition, the presence of the suggested 9 bp mt promoter of *Saccharomyces cerevisiae*, which is known as non-anucleotide (Osinga et al., 1982), along with the proposed mt promoter sequences of other fungal species, i.e., *Neurospora crassa* (Kennell and Lambowitz, 1989; Kubelik et al., 1990; Kleidon et al., 2003), *Schizosaccharomyces pombe* (Schäfer, 2005; Schäfer et al., 2005), *Candida albicans* (Kolondra et al., 2015), and *Starmellera bacillaris* (Pramateftaki et al., 2006) were searched manually as they were helpful for defining the putative promoter sequences of the mt

TABLE 1 | Oligonucleotide primers used for PCR and RT-PCR assays in this work.

Primer name	Primer sequence (5′3′)	Length (bp)	Tm (°C)	GC (%)	PCR product (bp)
nad1UF	5′-TATGCAAAGAAGAATAGGTCCAAAT-3′	25	52.6	32	560
nad1UR	5′-TAACCACCTAAGAATAAAATAGTAG-′	25	44.2	28	
nad1b	5′-GCATGTTCTGTCATAAASCCACTAAC-3′	26	53.1	42.3	
cox1UF	5′-CAATTATACAATGTTATAGTTACTGCTCAT-3′	30	51.0	26.7	600
cox1UR	5′-CCAAAACCAGGTATAATTACAATATA-3′	26	48.7	26.9	
cox2UF	5′-GCTTTCCCTTCATTTAATTTATTGTATTTA-3′	30	55	23.3	350
cox2UR	5′-AAGCAGATGCTTGATTTAATCTACCAG-3′	27	55.7	37	
cox3UF	5′-CATATAGTATCACCATCACCTTGACC-3′	26	52.8	42.3	460
cox3UR	5′-ACATGTAATCCATGGAAACCTG-3′	22	50.6	40.9	
cobUF	5′-GTTATTACTAATTTATTAAGTGCGTTCC-3′	29	50.2	27.6	400
cobUR	5′-GCATAGAAAGGTAGTAAGTATCATTC-3′	26	47.5	34.6	
rnsUF	5′-GCCAGCAGTCGCGGTAATAC-3′	20	54.8	60	600
rnsUR	5′-TATAAAGGCCATGATGTCTTGTCTT-3′	25	53.2	36	
atp6UF	5′-CATGGATTAGAATTCTTCTCATTA-3′	24	47.2	29.2	200
atp6UR	5′-ATATTAGCAGCTAAACTAAAACCTA-3′	25	46.3	28	
atp8UF	5′-ATGCCACAATTAGTACCATTTTA-3′	23	48.2	30.4	800
nad2PF	5′-TATTATTAATATCTAGTAGTGATTTAGTATC-3′	31	44.1	19.4	400
nad2PR	5′-ACAAAAGTTGTAACAATTGTAGGTAT-3′	26	48	26.9	
nad4PF	5′-GATGGTATATCTATATATTCTGTATTATTGAC-3′	32	48.9	25	737
nad4PR	5′-CCTTCTATACCTTGTATTGTATTACTAAA-3′	29	48.8	27.6	
nad4LF	5′-GAGGAAGTATAGCAATAGAGTATAAATAA-3′	29	48.3	27.6	400
nad5PoR	5′-CATCATAAAAGTAAATAAAACTAAATA-3′	27	44.5	14.8	
nad5PF	5′-TATTTAAGTTTATTTACTTTTATGATG-3′	27	44.5	44.5	500
nad5PR	5′-ACTAAATACTGTAGTTATAGCACCTAATC-3′	29	48.1	31	
nad6PF	5′-GTATATATAGGAGCTGTATCAATCTTA-3′	27	45.6	29.6	337
nad6PR	5′-CACCTACCATAGCTAATAATAGAAT-3′	25	45.9	32	
rps3F	5′-AAAAATATTCCAAGCCATCAT-3′	21	47.7	28	400
rps3R	5′-TTAGTGCCTTCAAAAACATTAT-′	22	46.1	27	
trnAF	5′-AGGTTCGATTCCTAGTTTCTCC-3′	22	50.4	44	400
trnQR	5′-GTGATTCGAACACCCACTATTG-3′	22	51.5	45	
trnTF	5′-GGCGCGATACCTCCATT-3′	17	50	58	1,200
trnM2R	5′-TGACTTGAACACTTATTATTACTG-3′	24	43	34	
atp8UR	5′-TTATATATTATTAATAAACATACGTGATAAG-3′	31	47.1	26	200
atp9F	5′-ATGTTACAATCTTCAAAAATAATAGG-3′	26	48	23	200
atp9R	5′-TTAAGCAACATTTAATAATAATAATG-3′	26	45.5	25	
rnloF	5′-CTAAGTTGGTTAAGGATAAGTG-3′	22	43	36	800
rnloR	5′-CATTTCTTTTCCTGAAGTTAC-3′	21	43.1	33	
nad4LFN	5′-GAATTTTAGGATTCGTATTTAATAG-3′	25	46	26	250
nad4LRN	5′-TTATTTATATTCTATTGAGATACTTCC-3′	27	45	28	
rnlF	5′-CAAAAGATATCAAAAGAGATTC-3′	22	42.6	27.8	600
rnlR	5′- CACTTATCCTTAACCAACTTAG-3′	22	43	36.4	
nad3F	5′- ATTTGAATGTGGTTTTCAT-3′	19	41.3	26.3	250
nad3R	5′- AATGCASTTTTACCTAATTCA-3′	21	44.3	28.6	
Rpo41F	5′-GCAAGATGCTCAGTCGACAAACAAGGAG-3′	28	64.5	48.3	4,300
Rpo41R	5′-GCTCAATAAAATACACCAAATGCCAACTC-3′	29	60.9	36.7	

genomes of these entomopathogenic fungi belonging to the order Hypocreales.

In silico Search for Termination Sequences of Transcription Units

To determine the presence of termination sequences for the mt transcription units, the nucleotide sequences downstream each mt transcription unit of both species were retrieved and aligned with Clustal Omega program (Sievers and Higgins, 2021). In comparison with the known yeast mt termination sequences (Osinga et al., 1984; Butow et al., 1989), no similar sequences were found. Thus, a comparative search among the different intergenic regions was performed to find the conserved termination sequences in the aligned matrices.

TABLE 2 | 5′RACE primers.

Oligo name	Oligo sequence (5′–3′)
Oligo d(T)-anchor primer	5′-GACCACGCGTATCGATGTCGAC TTTTTTTTTTTTTTTV-3′
PCR anchor primer	5′-GACCACGCGTATCGATGTCGAC-3′
NAD1.pr.ex.R	5′-CCATAATAACCTACAGCATTAGGCCCTAATCTTC-3′

Protein Molecular Modeling

The sequence of Rpo41 of *M. brunneum* was retrieved from the whole genome (Nucleotide Acc No. NW_014574670.1) and primers were designed at the 5′ and 3′ end of the gene (**Table 1**). The amplicon was cloned and sequenced in both directions in order to verify the sequence before analyzing its secondary structure. Prediction of the Rpo41 secondary structure was performed using the PSIPRED 4.0 Workbench (UCL-CS Bioinformatics, London, United Kingdom) (Buchan and Jones, 2019). For homology modeling of the Rpo41 protein, the Hidden Markov Model-based tool HHPred (Zimmermann et al., 2018) and MODELLER 9.25 (Webb and Sali, 2016) were used, based on the highly similar crystal structure of yeast mitochondrial RNA polymerase from *S. cerevisiae* (Protein Data Bank, PDB 6YMW) with 98.94% probability and 1.4×10^{-21} *E*-value, as described previously (Deshpande and Patel, 2012; De Wijngaert et al., 2021). Protein structures were visualized and compared using PyMOL 2.4.[1]

Polymerase Chain Reaction Amplification and Cloning

Mitochondrial gene fragments for seven mt genes (*rnl*, *rps*3, *nad*3, *cob*, *cox*1, *nad*4, and *cox*3) were amplified by PCR to produce probes for Northern hybridization analyses. Newly designed primers, as well as primers proposed by Kouvelis et al. (2004) and Ghikas et al. (2006) were used (**Table 1**) and 1 μl total DNA (1 μl) from each strain was used as template. Polymerase chain reaction (PCR) amplifications were performed with Taq DNA polymerase (Invitrogen, Waltham, MA), in a FastGene® Ultra Cycler Gradient (Nippon-Genetics Europe GmbH, Dueren, Germany), following previously described protocols (Kouvelis et al., 2004). Amplicons were separated in 1.5% agarose–TAE gels and amplicons' purification was performed afterward, using Monarch® PCR and DNA Cleanup Kit (New England Biolabs, Hitchin, United Kingdom), NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel, Dueren, Germany). The vector pGEM®-T was used along with the pGEM®-T Vector System (Promega, Madison, WI) for the cloning. All recombinant plasmids were screened to verify that the insert's size coincided with the amplicon's size and three randomly chosen recombinant plasmids were sequenced. All fragments were sequenced in both directions and since sequences were identical (100% id) with the those found in the complete genomes of both *B. bassiana* and *M. brunneum,* they were not submitted to any of the publicly available databanks like GenBank or ENA of EMBL-EBI.

[1]https://pymol.org

Reverse Transcription-Polymerase Chain Reaction

The isolated total RNA (1 μg) was reverse transcribed to generate single-stranded complementary DNA (cDNA) using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) with reduced RNase H activity (Conc. 200 U/μl, SuperScript™ II Reverse Transcriptase, Invitrogen, Waltham, MA), according to the manufacturer's instructions. For the analysis of multiple target mitochondrial RNAs, each first-strand cDNA was synthesized by priming with a gene-specific primer (**Table 1**). The cDNAs were stored at −80°C. After reverse transcription (RT), the cDNAs from different target RNAs were used as a template for amplification with specific primers by PCR (KAPA HiFi HotStart ReadyMix PCR Kit, Merck KGaA, Darmstadt, Germany). A negative control (DNAse-treated RNA, no reverse transcription) did not produce an amplification product in the PCR, confirming the cDNA (and not the mitochondrial DNA) origin of the PCR product. DNA fragments generated with KAPA HiFi HotStart DNA Polymerase (15 s extension per cycle for targets $\leq$ 1 kb, and 45 s/kb for longer fragments) have been used directly for blunt-end cloning. PCR products were cloned in vector pBluescript II SK (Stratagene, Agilent Technologies, Santa Clara, CA), subcloned as smaller fragments to pTZ57R/T cloning vector (only for large pcr fragments $>$ 2 kb) and sequenced in both directions using the M13 Forward and Reverse universal primers. Amplicon sequences were analyzed using the «Sequence Scanner v.2» software (Thermo Fisher Scientific Inc., Waltham, MA). Additionally, the RT-PCR amplicons were separated on a 1,2% (w/v) agarose gel containing ethidium bromide and visualized under UV light (UVP Gel Imaging System, Thermo Fisher Scientific Inc., Waltham, MA).

Northern Hybridizations

For Northern hybridization analysis, total cellular RNA (30 μg/lane) was electrophoresed in 1.5% formaldehyde-agarose gels and transferred to positively charged nylon filters. Specifically, RNA were resuspended in 50% deionized formamide, 10 mM morpholinepropanesulfonic acid (MOPS) pH 7.0, 1.4 mM sodium acetate, 0.5 mM EDTA, and 2.2 M formaldehyde. The RNA was denatured at 65°C for 15 min and fractionated on 1.5% agarose gel prepared in 1x formaldehyde gel running buffer (FGRB) with 2.2 mol/l formaldehyde, based to the protocol provided by DIG-High Prime DNA Labeling and Detection Starter Kit I (Roche Diagnostics, Sigma-Aldrich Inc., Darmstadt, Germany). The integrity of the electrophoresed total RNA was monitored by ethidium bromide staining of formaldehyde-agarose gel. The separated RNA fragments were then blotted onto a nylon filter (Porablot™ NY plus, Macherey-Nagel, Dueren, Germany) with 20x SSC (3 mol/l sodium chloride and 0.3 mol/l sodium citrate, pH 7.0) by applying a low pressure vacuum (VacuGene XL Vacuum Blotting System, Amersham Biosciences Corp., Buckinghamshire, United Kingdom). The filter was allowed to dry at room temperature for at least 30 min and the blotted RNA was immobilized on the filter by UV cross-linking. DIG-labeled probes were used for

TABLE 3 | Distribution of transcription factors and RNA polymerase subunits used in the phylogenetic analyses from species belonging to all domains of life and their taxonomic classification.

	Taxonomy					
	Domain	**Kingdom**	**Phylum**	**Subphylum**	**Class**	**Number of strains**
T3/T7 RNAP	Viruses	Heunggongvirae	Uroviricota	–	Caudoviricetes	2
aTBP/rpoB	Archaea	Crenarchaeota	Crenarchaeota	–	Thermoprotei	45
		Euryarchaeota	Hadesarchaeota	–	Hadesarchaea	3
		Proteoarchaeota	Verstraetearchaeota	–	–	3
			Odinarchaeota			1
			Thorarchaeota			1
			Lokiarchaeota;			3
σ-factor/rpb2	Bacteria	–	Proteobacteria	–	Alphaproteobacteria	17
					Betaproteobacteria	12
					Gammaproteobacteria	8
		Terrabacteria	Firmicutes	–	Bacilli	4
					Clostridia	6
			Cyanobacteria—Melainabacteria	Cyanobacteria	Cyanophyceae	13
			Deinococcus-Thermus	–	Deinococci	6
mtf1/rpo41	Eukaryota	Fungi	Ascomycota	Pezizomycotina	Dothideomycetes	10/0/0/0
					Eurotiomycetes	9/0/0/0
					Leotiomycetes	9/0/0/0
					Sordariomycetes	10/2/0/0
TBP(SL1)-RPA2				Saccharomycotina	Saccharomycetes	3/0/22/25
				Taphrinomycotina	Pneumocystidomycetes	4/0/3/0
					Schizosaccharomycetes	5/0/7/3
			Basidiomycota	Agaricomycotina	Agaricomycetes	16/0/0/0
TBP-RPB2				Pucciniomycotina	Microbotryomycetes	1/0/0/0
				Ustilaginomycotina	Ustilaginomycetes	1/0/0/0
			Early diverging fungal lineages	Mucoromycotina	Incertae cedis	3/0/0/0
TBP-RPC2		Metazoa	Chordata	–	Mammalia (Mammals)	2/41/21/25
				–	Reptiles	0/1/9/1
				–	Amphibia	0/29/0/1

hybridization to blotting membranes as defined in standard methods (DIG-High Prime DNA Labeling and Detection Starter Kit I, Roche Diagnostics, Sigma-Aldrich Inc., Darmstadt, Germany). Mitochondrial DIG-labeled DNA probes (rnl, rps3, nad3, cob, cox1, nad4, cox3) were generated with DIG-High Prime solution according to the random primed labeling technique. Each northern blot was pre-hybridized with an appropriate volume (10 ml/100 cm^2 filter) of DIG Easy Hyb buffer (5x SSC, 50% formamide deionized, 0,1% sodium-lauroylsarcosine, 0.02% SDS, and 2% Blocking Reagent) to hybridization temperature for 30 min with gentle agitation in an appropriate container (Hybaid H-9360 Hybridization Oven, Marshall Scientific LLC., Hampton, NH). Denatured DIG-labeled DNA probe were added to pre-heated DIG Easy Hyb buffer and hybridization was performed for 16hrs at 65°C (RNA). The RNA blots were washed with 2x SSC, 0.1% SDS for 25 min at room temperature under constant agitation. Subsequently, membranes were used directly for detection of mitochondrial polycistronic transcripts. The hybridized DNA probes were immunodetected with anti-digoxigenin-AP (Fab fragments) and were then visualized with the colorimetric substrates NBT/BCIP. Immuno—colorimetric detection was performed as described by DIG DNA Detection Kit. Results of Northern Blot Analysis were documented by photocopying the wet filter (BioSpectrum® Imaging System, UVP Gel Imaging System, Thermo Fisher Scientific Inc., Waltham, MA).

RACE Polymerase Chain Reaction

The 5′ end UTR of mt transcripts of a single polycistronic transcript has been determined using 5′-RACE (see section "Results"). Lyophilized, ground mycelium (100 mg) of the wild-type strains *B. bassiana* and *M. brunneum* was used for extraction of total RNA, as previously described. DNA was removed from the RNA sample by treatment with DNase I (Takara Bio Inc., Shiga, Japan). Reverse transcription and PCR reactions were carried out using a 5'/3' RACE Kit (2nd Generation, Roche Applied Science, Merck KGaA, Darmstadt, Germany). The sequence of PCR primers used for definition of the 5′-ends of all mitochondrial transcripts is given in **Table 2**. Amplicons were purified and sequenced as described

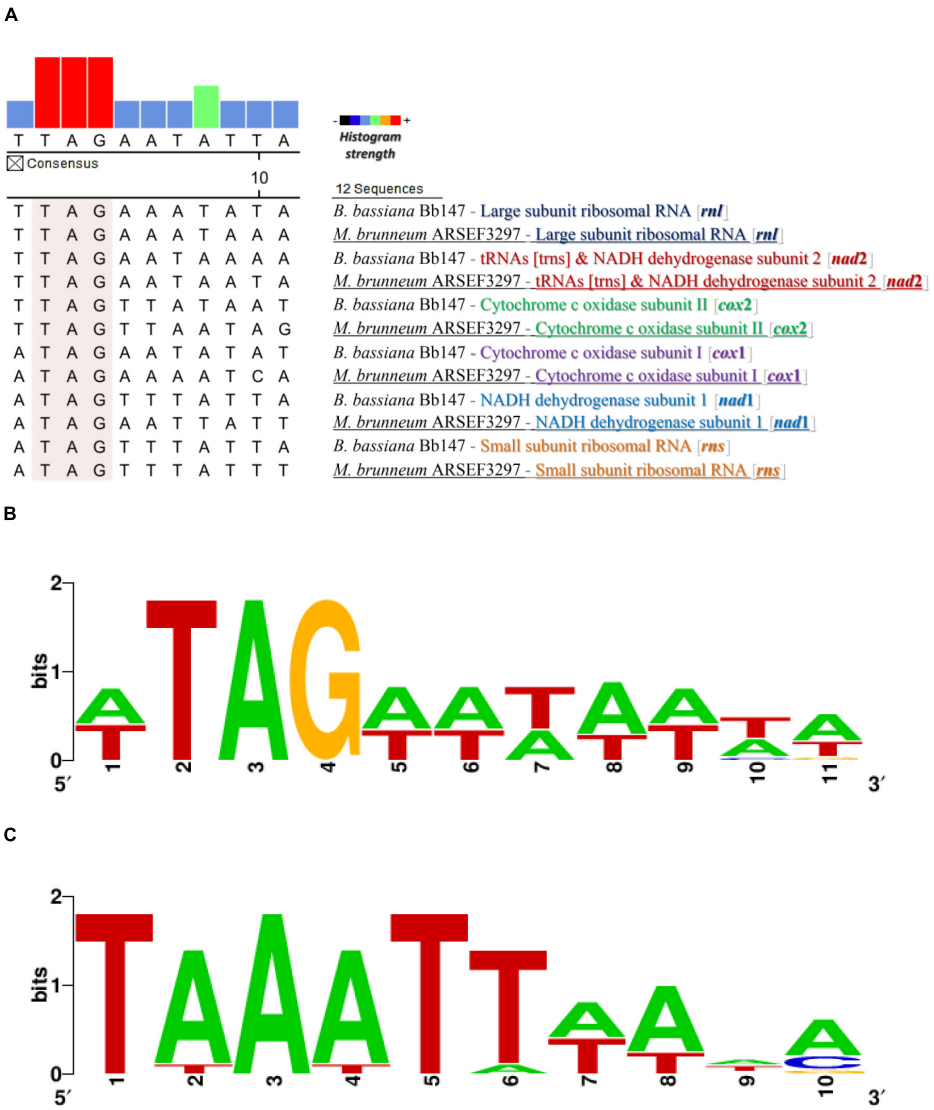


FIGURE 1 | The promoter and transcription termination sequences as found from the *in silico* analyses. **(A)** Putative promoter sequence alignment for different mt genes from *M. brunneum* ARSEF3297 and *B. bassiana* Bb147. **(B)** The motif of the putative mt promoter produced by this alignment analysis. **(C)** The distinct putative mt transcription termination motifs that are located at a variable distance (1–218 nt) downstream from mRNA and rRNA coding regions in the genera *Metarhizium* and *Beauveria*.

previously. The 5′ end UTR sequences of primary mitochondrial polycistronic transcripts have been determined *in silico* (see section "*In silico* Prediction of Promoters"), since the sequences were identical to the respective genomic fragments of the complete mt genomes of *M. brunneum* ARSEF 3297 and *B. bassiana* Bb 147 (GenBank Acc Nos: NW_014574670.1 and EU100742.1, respectively).

Phylogenetic Analyses and Methods of Examining Co-evolution

Phylogenetic analyses of mitochondrial Rpo41 and Mtf1 homologs were performed with the inclusion of sequences from four different taxonomic domains (Viruses, Archaea, Bacteria and Eukaryota; **Table 3** and **Supplementary Table 3**). Phylogenetic analyses of Rpo41 and Mtf1 proteins were done separately for each protein at first, and then the matrices were concatenated to a single matrix. The respective proteins from other species were collected after a BLASTp search and all protein sequences were aligned with Clustal Omega (EMBL-EBI). The alignments were transferred to the program PAUP 4.0b10 (Swofford, 2002) for the Neighbor Joining (NJ) method, keeping all the parameters at default values. The topologies of the obtained trees were assessed by the bootstrap method (NJ-BP) and the number of replicates was set to 1,000. Each NJ tree was presented using FigTree v1.4.3.[2] The maximum likelihood (ML) method was further employed to construct

[2]http://tree.bio.ed.ac.uk/software/figtree/

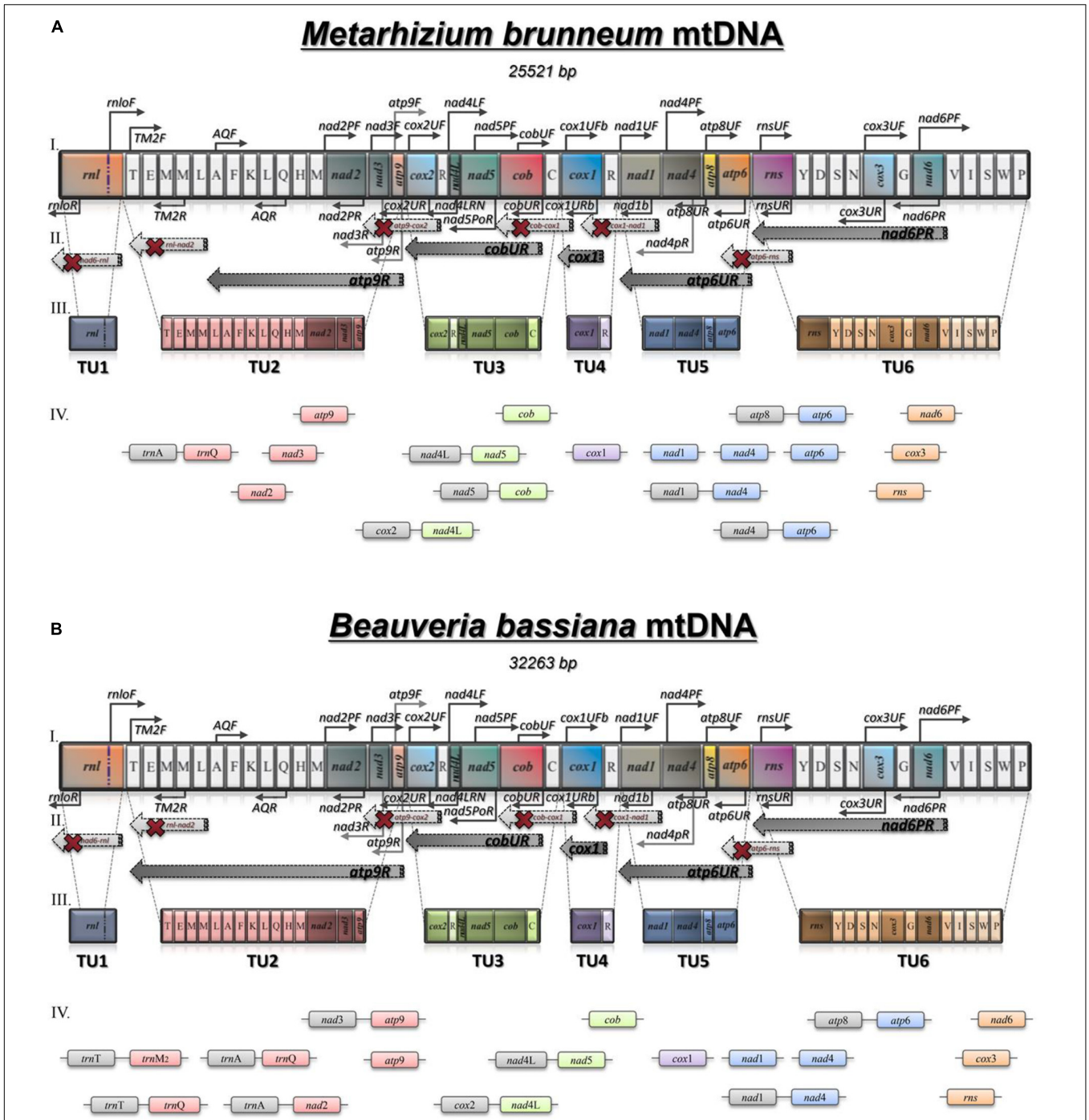


FIGURE 2 | The mt transcripts of *M. brunneum* **(A)** and *B. bassiana* **(B)**: Schematic representation of the strategy followed to determine transcripts for both species: (i) The synteny of the mt genome. Black arrows above and below the genes of the mt genome show the primers used for the RT- PCR experiments. (ii) Dark gray arrows represent the successfully obtained RT-PCR amplicons, while light gray arrows with a red "X" represent RT-PCR efforts with no result. (iii) The block of genes included in primary transcripts as acquired by Northern hybridizations. (iv) The amplicons obtained with RT-PCR.

trees using RaxML (version 8.2.9) (Stamatakis, 2014). The best protein substitution model was found using ProtTest v3 (Darriba et al., 2011). For the matrices of RNA Polymerases and their transcription factors, the WAG + I + G + F model was found as most suitable, while the concatenated matrix had the LG + G model. Alpha values were set to 2.993, 2.694, and 2.959 for the RNA polymerases, Transcription Factors and concatenated matrices, respectively. All other parameters were estimated by RaxML. Results were assessed using 100 bootstrap replicates.

In order to further examine the putatitve co-evolution of Rpo41 and Mtf1, three different methods were applied, STRING (Szklarczyk et al., 2021), Mirror Tree (Ochoa and Pazos, 2010), and Evolutionary Trace (Morgan et al., 2006). For both methods, STRING and ETviewer servers were used.[3,4] In STRING method, the single proteins of *M. brunneum* and *S. cerevisiae* were used as input, under default parameters, while in the second method, Rpo41 (PDB: 6YMW) and Mtf1 (PDB: 1I4W) of *S. cerevisiae* were used. For the Mirror Tree, the method was employed through the server[5] using as input the amino acid sequences from yeast. The comparison of the trees produced was impossible through the same server and thus, the single trees were obtained in Newick format and then compared with the phylo.io program (Robinson et al., 2016).[6]

RESULTS

Mitochondrial Genome Annotation and Analysis

The complete mt genomes of both species, i.e., *B. bassiana* and *M. brunneum* used in this work are publicly available (Ghikas et al., 2010; Kortsinoglou et al., 2020). In this work, the strains used were ATHUM 4946 and ARSEF 3297, respectively. Since the whole genome sequence of the *M. brunneum* ARSEF 3297 is available, the mtDNA was retrieved and annotated by comparison to the known and annotated mt genome of strain ARSEF 4556 (Kortsinoglou et al., 2020) as described in "Materials and Methods". Therefore, the mtDNA of this *M. brunneum* strain was found to be a circular molecule of 25,521 bp (**Supplementary Figure 1** and **Supplementary Table 1**), with an overall G + C content of 28.8%. It contained all 14 protein encoding genes found in mtDNAs of the Pezizomycotina (Ghikas et al., 2006), the genes encoding for the large and small subunits of the ribosome (*rnl* and *rns*), and 25 genes encoding for tRNAs. A putative ribosomal protein (*rps*3) was also detected within the large ribosomal subunit using the *rps*3 homologs motif in fungal genomes (Bullerwell et al., 2000; Korovesi et al., 2018). In general, this mitogenome showed to be highly conserved to the genome of ARSEF 4556. Specifically, synteny was identical and the only variability could be observed in three intergenic regions. The ARSEF 3297 and ARSEF 4556 strains presented mt intergenic regions of sizes 4,069 and 4,215 bp, respectively. The largest intergenic region was that of *rns*-*trn*Y (496 bp) and the smallest was the 1bp overlap of genes *nad*4L-*nad*5. The identity of the intergenic sequences for both strains was 100% with the exception of the sequences that differentiated the size of the three intergenic regions, i.e., *nad*5-*cob*, *atp*6-*rns*, and *rns*-*trn*Y (with the addition of 198, 452, and 496 bp in ARSEF 3297).

[3]https://string-db.org/cgi/input?sessionId=buxdn3JsL4Jy&input_page_show_search=on

[4]http://evolution.lichtargelab.org/ETviewer

[5]http://csbg.cnb.csic.es/mtserver/#

[6]http://phylo.io/index.html

In silico Prediction of Promoters

The complete mt genomes of the species used in this work were searched for the existence of putative promoter sequences along with other known Hypocrealean mitogenomes based on known promoters from other fungal species as described previously in "Materials and Methods." More emphasis was given to the intergenic regions of the complete mt genomes. It is worth mentioning that the intergenic regions were found to be highly enriched in A/T sequences. More specifically, the intergenic regions of the mt genomes of *B. bassiana* and *M. brunneum* presented A/T content of 73 and 76%, compared to the complete genomes' A/T content of 72 and 71%, respectively. The search of possible repetitive elements did not present any conserved elements (data not shown), making, thus, the *in silico* identification of putative promoters extremely difficult. However, the consensus putative promoter sequence obtained by these comparative analyses was determined to be TTAGAATATTAT on the mitochondrial genomes of *Beauveria* and *Metarhizium* species (WTAGWWHWWHD: a modified consensus mt-promoter sequence within the order Hypocreales—data upon request) and this sequence was found scattered in six (6) intergenic regions (**Figure 1**).

3′ End Processing of Mitochondrial Transcripts

Most of the tRNA genes in *M. brunneum* ARSEF3297 and *B. bassiana* ATHUM4946 mitochondria are present as two major clusters upstream and downstream from the *rnl* gene, and are co-transcribed with mRNA sequences, according to the Northern Hybridizations. In at least two cases, initial transcripts (*cox*2-*trn*R1-*nad*4L-*nad*5-*cob*-*trn*C and *cox*1-*trn*R2) include tRNA sequences and are subsequently processed to generate the mature RNAs. The endpoints of these abundant mitochondrial transcripts generally coincide with those of tRNA sequences. We therefore conclude that tRNA sequences in some polycistronic transcripts act as primary signals for RNA processing in mitochondria of these entomopathogenic fungi (Punctuation model—single tRNAs may play this role). The situation is somewhat analogous to that observed in mammalian mitochondrial systems (Ojala et al., 1981; Rorbach and Minczuk, 2012) and in other filamentous fungi (*N. crassa*) (Breitenberger et al., 1985; Burger et al., 1985). In this study, the strongest evidence for tRNA sequences being primary signals in mitochondrial RNA processing comes from an analysis of transcription units 4 and 5 (*cox*1-*trn*R2 and *nad*1-*nad*4-*atp*8-*atp*6, respectively). Experimental results obtained by determining the 5′ end of the polycistronic transcript *nad*1-*nad*4-*atp*8-*atp*6 (5′ end RACE-PCR approach) indicated that *trn*R2 is co-transcribed with the gene *cox*1 (supported also by the Northern hybridization experiments). Furthermore, the 3′-termini of mRNAs and LSU rRNA are proposed to be distinct 5′-TAAATT-3′ motifs that are located at a variable distance (1–218 nt) downstream from mRNA and LSU-rRNA coding regions. Similarly, 3′-RNA processing motifs are also present in budding yeasts that have functionally analogous A + T rich

TABLE 4 | Pre-mature polycistronic transcripts and their sizes as produced by the RT-PCR experiments for both *B. bassiana* and *M.* anisopliae.

A/A	Gene composition of the polycistronic transcripts	Size (Kb)	
		Bb*	Mb*
I	***rnl*** (***rps3***)	∼5	∼4.7
II	*trn*T-*trn*E-*trn*M1-*trn*M2-*trn*L1-*trn*A-*trn*F-*trn*K-*trn*L2-*trn*Q-*trn*H-*trn*M3-*nad*2-***nad3***-*atp*9	∼4.5	∼3.5
III	*cox*2-*trn*R1-*nad*4L-*nad*5-***cob***- *trn*C	∼6	∼5
IV	***cox1***- *trn*R2	∼4.5	∼2
V	*nad*1-***nad4***-*atp*8-*atp*6	∼6	∼4
VI	*rns*-*trn*Y-*trn*D-*trn*S-*trn*N-***cox3***-*trn*G-*nad*6-*trn*V(*trn*L)-*trn*I-*trn*S2-*trn*W-*trn*P	∼5 kb	∼5.5

*The mt genes or parts of the genes of M. brunneum which have been used as probes in the Northern hybridization experiments in bold. *Mb, M. brunneum ARSEF3297, Bb, B. bassiana ATHUM4946.*

TABLE 5 | Amplicons produced with RT-PCR experiments, and their sizes after the PCR amplification in kbs.

First-Strand cDNA Synthesis (Reverse primer)	Reverse transcription PCR (Primer Pair)	Gene or region amplified after 2nd PCR/length (kb)	Bb amplicon size (kb)	Mb amplicon size (kb)
cox1URb	cox1UFb-cox1URb	*cox*1	1	1
	cobUF-cox1URb	*cob*-*cox*1	ND	ND
cobUR	cobUF-cobUR	*cob*	0.45	0.45
	nad4LF-nad5PoR	*nad*4L-*nad*5	0.4	0.4
	nad5PF-cobUR	*nad*5-*cob*	NT	2.1
	nad5PF-nad5PR	nad5	0.5	0.5
cox2UR	nad3F-cox2UR	*nad*3-*cox*2	ND	ND
	atp9F-cox2UR	*atp*9-*cox*2	ND	ND
nad4LRN	atp9F-nad4LRN	*atp*9-*nad*4L	ND	ND
	cox2UF-nad4LRN	*cox*2-*nad*4L	1.1	1.1
nad2PR	TM2F-TM2R	*trn*T-*trn*M2	1.2	NT
	TM2F-AQR	*trn*T-*trn*Q	1.7	NT
	AQF-AQR	*trn*A-*trn*Q	0.5	0.5
	AQF-nad2PR	*trn*A-*nad*2	1.5	NT
	rnloF-nad2PR	*rnl*-*nad*2	ND	ND
atp9R	atp9F-atp9R	*atp*9	0.2	0.2
	nad2PF-nad2PR	*nad*2	NT	0.4
	nad3A-nad3B	*nad*3	NT	0.3
	nad3A-atp9R	*nad*3-*atp*9	0.75	NT
rnloR	cox3UF-rnloR	*cox*3-*rnl*	ND	ND
	nad6PF-rnloR	*nad*6-*rnl*	ND	ND
nad6PR	rnsUF-rnsUR	*rns*	0.6	0.6
	cox3UF-cox3UR	*cox*3	0.45	0.45
	nad6PF-nad6pR	*nad*6	0.4	0.4
rnsUR	nad4PF-rnsUR	*nad*4-*rns*	ND	ND
	atp8UF-rnsUR	*atp*8-*rns*	ND	ND
atp6UR	atp8UF-atp6UR	*atp*8-*atp*6	0.9	0.85
	atp6UF-atp6UR	*atp*6	0.25	0.25
	nad1UF-nad4PR	*nad*1-*nad*4	2.4	2.2
	nad1UF-nad1UR	*nad*1	0.55	0.55
	nad4PF-nad4PR	*nad*4	0.7	0.7
	nad4PF-atp6UR	*nad*4-*atp*6	NT	2
atp8UR	cox1UF-nad1UR	*cox*1-*nad*1	ND	ND
nad1b	cox1UFb-nad1b	*cox*1-*nad*1	ND	ND

Bb and Mb denote B. bassiana and M. brunneum, respectively. ND, not determined; NT, non-tested.

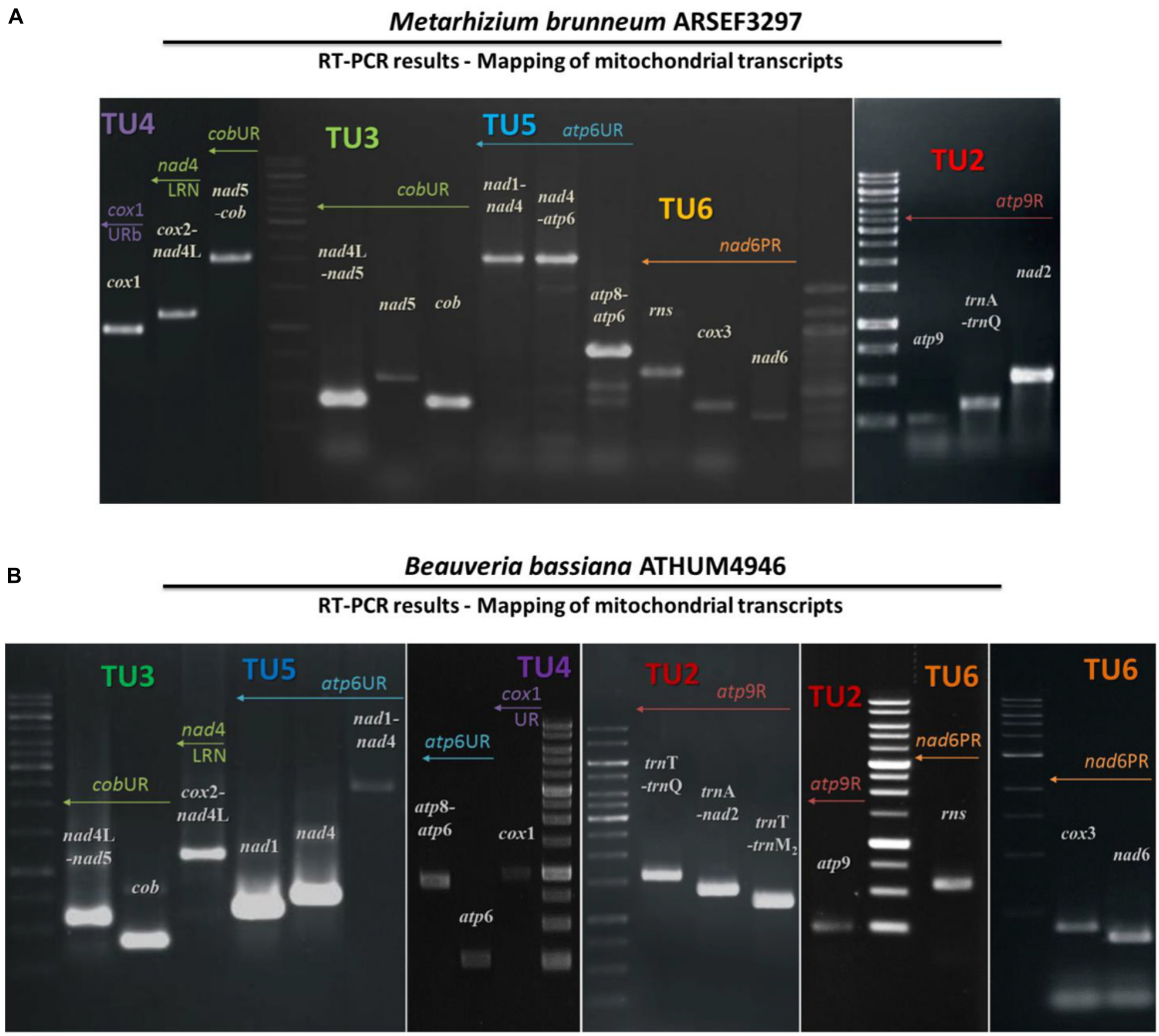


FIGURE 3 | The amplicons of mitochondrial primary transcripts as obtained by RT-PCR followed by PCR in the entomopathogenic fungi *M. brunneum* ARSEF3297 **(A)** and *B. bassiana* ATHUM 4946 **(B)**. The actual amplicons obtained. Ladder 1kb (GeneRuler 1 kb DNA Ladder, Thermo Fisher Scientific Inc, Waltham, MA) was used as marker. TU, transcription unit.

dodecamer processing signals (Osinga et al., 1984; Butow et al., 1989; Hofmann et al., 1993).

Defining the Polycistronic Transcripts

An RT-PCR strategy was developed to determine the number and content of premature polycistronic transcripts. The synteny of both mitogenomes of *M. brunneum* and *B. bassiana* are almost identical (with the differences located on the sequences of their intergenic regions; **Figure 2**). Therefore, the same sets of primers were designed and used for both fungal species, as described in "Materials and Methods."

The amplified polycistronic transcripts were produced after amplification of cDNAs using pairs of primers for neighboring genes. This strategy, showing all combinations of primers (both successful and failed ones) in order to determine the polycistronic units is shown in **Figure 2**. After the completion of all RT-PCR experiments it became evident that there were 6 major premature polycistronic transcripts as shown in **Table 4**.

Based on these RT-PCR results from both entomopathogenic fungi, it is worth mentioning that the cDNA generated with primer cox1URb produced a monocistronic transcript for *cox*1 gene, but it is possible that this gene can be co-transcribed with *trn*R2 as shown from the Northern hybridization experiments and the expected by size band (**Supplementary Figure 2**). The effort to amplify the *cob-cox*1 region did not produce any amplicon, indicating that the transcription of these two genes occurs in different polycistronic molecules. However, *cob* is co-transcribed with *nad*5, as shown in the case of *M. brunneum*, where a 2.1 Kb *cob-nad*5 amplicon is retrieved from the cDNA of cobUR (**Table 5** and **Figures 2**, **3**) and the amplification of both *nad*5 and *nad*4L-*nad*5 regions are successful for both species (**Table 5** and **Figures 2**, **3**). Moreover, neither *nad*3 nor *atp*9 gene could be identified from cDNA generated by

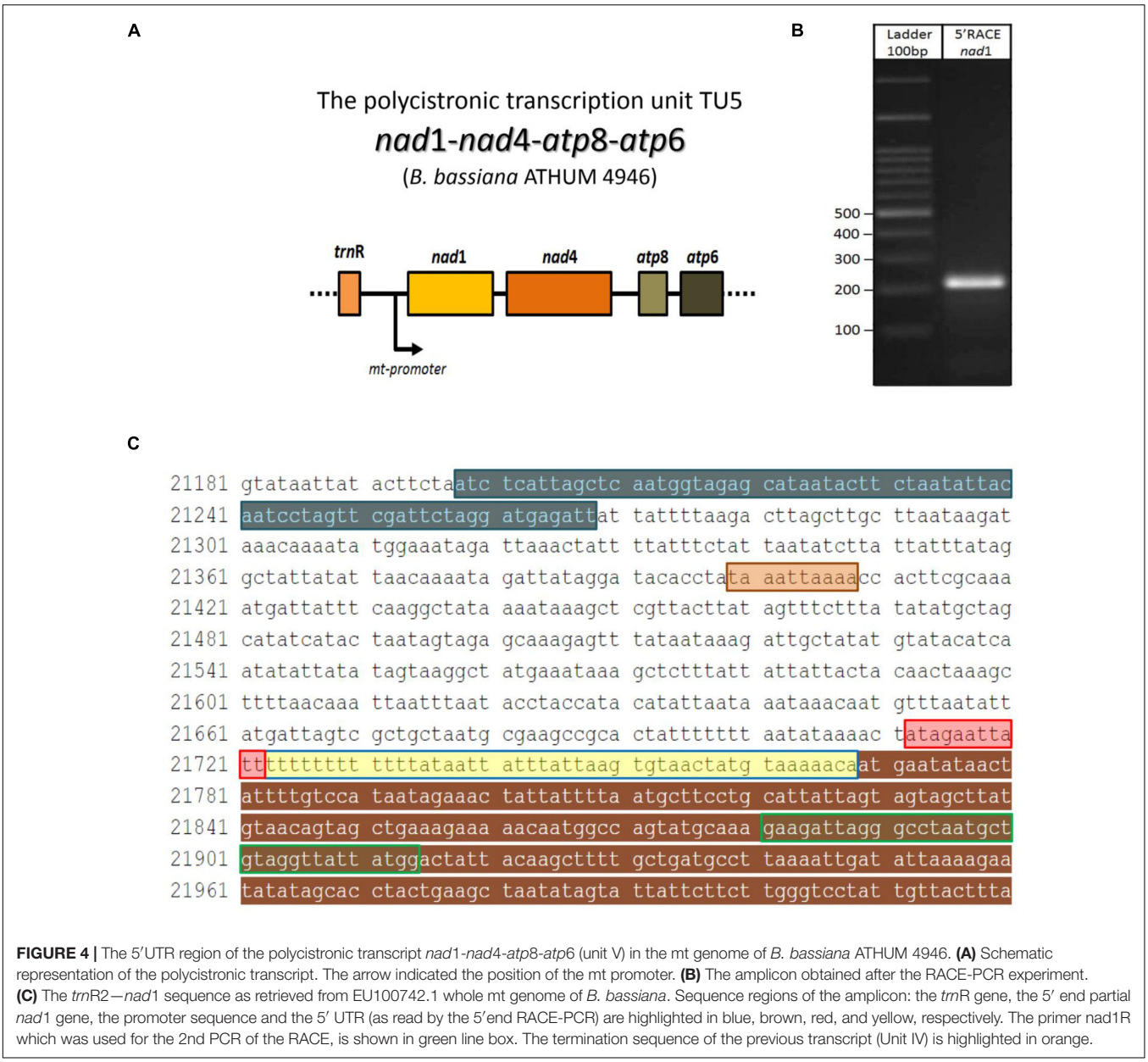


FIGURE 4 | The 5′UTR region of the polycistronic transcript *nad*1-*nad*4-*atp*8-*atp*6 (unit V) in the mt genome of *B. bassiana* ATHUM 4946. **(A)** Schematic representation of the polycistronic transcript. The arrow indicated the position of the mt promoter. **(B)** The amplicon obtained after the RACE-PCR experiment. **(C)** The *trn*R2—*nad*1 sequence as retrieved from EU100742.1 whole mt genome of *B. bassiana*. Sequence regions of the amplicon: the *trn*R gene, the 5′ end partial *nad*1 gene, the promoter sequence and the 5′ UTR (as read by the 5′end RACE-PCR) are highlighted in blue, brown, red, and yellow, respectively. The primer nad1R which was used for the 2nd PCR of the RACE, is shown in green line box. The termination sequence of the previous transcript (Unit IV) is highlighted in orange.

primers nad4LRN and cox2UR, suggesting that this set of genes is transcribed separately from the *cox*2-*trn*R-*nad*4L-*nad*5-*cob* primary transcript (**Figure 2**). Additionally, by using either primers atp9R or nad2PR for first strand cDNA, followed by PCR with primers amplifying *nad*2, *nad*3, *atp*9, and the regions *trn*T-*trn*Q (which contains tRNA genes for amino acids $TEM_1M_2L_1AFKL_2Q$) and *nad*3-*atp*9, amplicons were obtained (**Table 5** and **Figure 3**). However, the region *rnl*-*nad*2 was not amplified, suggesting the existence of two different transcripts for these two regions (**Figure 2**). This implies that the *trn* gene cluster as well as genes *nad*2, *nad*3, *atp*9 are co-transcribed as a pre-mature unit in both strains (**Figure 2**). Solely based on RT-PCR, the tRNAs lying upstream of the above transcription unit are co-transcribed with genes *nad*2, *nad*3, and *atp*9. No PCR products were obtained with the primer pairs nad6PF/rnloR and cox3UF/rnloR with cDNA generated with RT primer rnloR (**Table 5** and **Figure 3**). Considering the above-mentioned RT-PCR results, it became clear that *cox*3 and *nad*6 are transcribed separately from the *rnl*-*rps*3 transcription unit, which was further verified when the cDNA template of nad6PR was used successfully for the amplification of *rns*, *nad*6, and *cox*3 genes (**Table 5** and **Figure 3**). Thus, *rns* and *cox*3 are co-transcribed with the adjacent *nad*6 gene. In view of all these data, co-transcription of *rns*-*trn*Y-*trn*D-*trn*S-*trn*N-*cox*3-*trn*G-*nad*6 was found. No PCR amplicon of either *atp*8 or *atp*6 was obtained from cDNA generated by primer rnsUR (**Table 5** and **Figure 2**), showing the existence of two different transcripts for *rns* and *atp*8-*atp*6. The co-transcription of *atp*8 with *atp*6 was confirmed

with the partial PCR amplification of the genes included in this cDNA, and with the amplification of the whole targeted sequence, when the cDNA of atp6UR was used as template. Similarly, based on the same cDNA, the corresponding single gene amplicons of *nad*1 and *nad*4 genes, as well as, the *nad*1-*nad*4 intergenic sequence were produced (**Table 5** and **Figures 2**, **3**). Finally, no product for the sequence between *cox*1-*nad*1genes was obtained from cDNA generated by primer nad1b (**Table 5** and **Figure 2**). Thus, *cox*1 is transcribed to a different molecule than the polycistronic transcript which contains *nad*1-*nad*4-*atp*8-*atp*6 (**Figure 2**). Therefore, the RT experiments, showed the composition of all polycistronic transcripts with the exception of the five tRNA genes, i.e., *trn*V, *trn*I, *trn*S, *trn*W, *trn*P. For that reason, their inclusion to a polycistronic transcript was found by Northern blot hybridization analysis, while all other RT based results were further verified with Northern hybridizations (**Table 4**).

DIG-labeled DNA probes from mitochondrial gene fragments of *rnl*, *rps*3, *nad*3, *cob*, *cox*1, *nad*4, and *cox*3 were used for verifying the pre-mature transcripts (**Table 4**). In all cases, several bands were hybridized with the probes in both species (**Supplementary Figure 2**). However, the largest hybridized band corresponded to the expected size of the primary polycistronic transcripts, and the smaller bands may denote maturation products (**Supplementary Figure 2**), according to the mechanisms proposed for *S. cerevisiae*, *Neurospora crassa,* and *Schizosaccharomyces pombe* (Kennell and Lambowitz, 1989; Kubelik et al., 1990; Dieckmann and Staples, 1994; Schäfer, 2005).

Determining the 5′ End of the Polycistronic Transcript *nad*1-*nad*4-*atp*8-*atp*6

Since the *in silico* putative promoters were further verified with the determination of the polycistronic transcripts, the approach of 5′ end RACE-PCR with subsequent sequencing of the obtained amplicon was selected in the case of transcript V (**Table 2**), as a sampling test to support further and combine both *in silico* and *in vitro* findings. This experiment showed that the 5′ end UTR of an mt transcript in *B. bassiana* is 46 bp upstream the ATG, which encodes the methionine of *nad*1 and is located immediately after the promoter of the polycistronic pre-mature transcriptional unit (**Figure 4**).

The Mitochondrial Transcription RNA Polymerase Rpo41 and Its Transcription Factor MTF1

All the above transcriptional elements and features, i.e., promoters, maturation signals, and primary polycistronic transcripts, are of utmost importance for the harmonious function of the mt RNA polymerase and its transcription factor MTF1. Both proteins are encoded by genes located in the nucleus of the species examined. Mt RNA polymerase (Rpo41) was found in contigs PPTI01000001.1 and AZNG01000005.1 of the genomes of *B. bassiana* and *M. brunneum* species (WGS projects: PPTI01 and AZNG01), respectively. Based on these sequences, primers were designed in order to amplify the whole gene in both species (**Figure 5**). Their sequences were identical to the corresponding gene sequences found in the WGS and thus, it was possible in all cases to identify (a) the mitochondrial target sequence and (b) the conserved functional regions of these proteins (**Figure 5A**). Especially in the case of Rpo41 of *M. brunneum* ARSEF 3297, the tertiary structure of the protein was built, and all expected conserved regions (**Figure 5C**) were identified (**Figure 5D**). The phylogenetic analysis performed in this work including representative sequences of RNA polymerases from all living organisms, showed that the Rpo41 proteins of the Hypocrealean entomopathogenic species cluster with their mitochondrial homologs from Pezizomycotina, and subsequently with other fungal subphyla and then with their counterparts from Metazoa (**Figure 6**). Their ancestral homolog seems to be the RNA polymerases found at phages and most specifically T7 or T3 phages, which are also related to the bacterial RNA polymerase subunit b (**Figure 6**). These enzymes are the ancestors of archaeal RpoB and all eukaryotic nuclear RNA polymerases like RPA2, RPB2, and RPC2 (**Figure 6**).

Since the phylogenetic topologies of Rpo41 with the rest RNA polymerases showed a preferred but still relaxed association with their bacterial counterpart (**Figure 6**), it became evident that the phylogenetic history of the only other needed protein for a functional Rpo41, i.e., Mtf1, had to be studied. It is known that Mtf1 protein assembles with the mitochondrial polymerase in order the latter to be functional and moreover, Mtf1 is similar to the bacterial homolog of sigma factor (Cliften et al., 2000). However, the obtained phylogeny (**Figure 7**) indicated that MTF1 showed relationship with Transcription Binding Protein (TFIII subunit B) and only basally to the RpoD of bacteria.

Moreover, since the evolution of both Rpo41 and Mtf1 proteins was under investigation, a concatenated matrix of these proteins and their nuclear and prokaryotic counterparts from representative species was created, resulting to a phylogenetic tree (**Figure 8**). This tree showed that their positioning was basal along with the T7/T3 phage RNA polymerases.

Furthermore, both Rpo41 and Mtf1 of both species examined in this work had all important amino acids which play role to their activity but also conformation conserved, the based on the analysis from ET-viewer (**Supplementary Table 4**). In addition, the network relationships of each protein examined under the STRING method showed always the greatest hit with its complementary, i.e., Rpo41 vs. Mtf1 and vice versa, irrelevant to the organism from which they were acquired (**Supplementary Table 5**). Finally, the last method applied, Mirror Tree showed that both trees present almost identical topologies when compared (**Supplementary Figure 6**). All methods provided strong indications that these two proteins co-evolved.

DISCUSSION

In this study an effort was made to decipher the mitochondrial transcription of the Hypocrealean entomopathogenic species *B. bassiana* and *M. brunneum*, and correlate any insights acquired

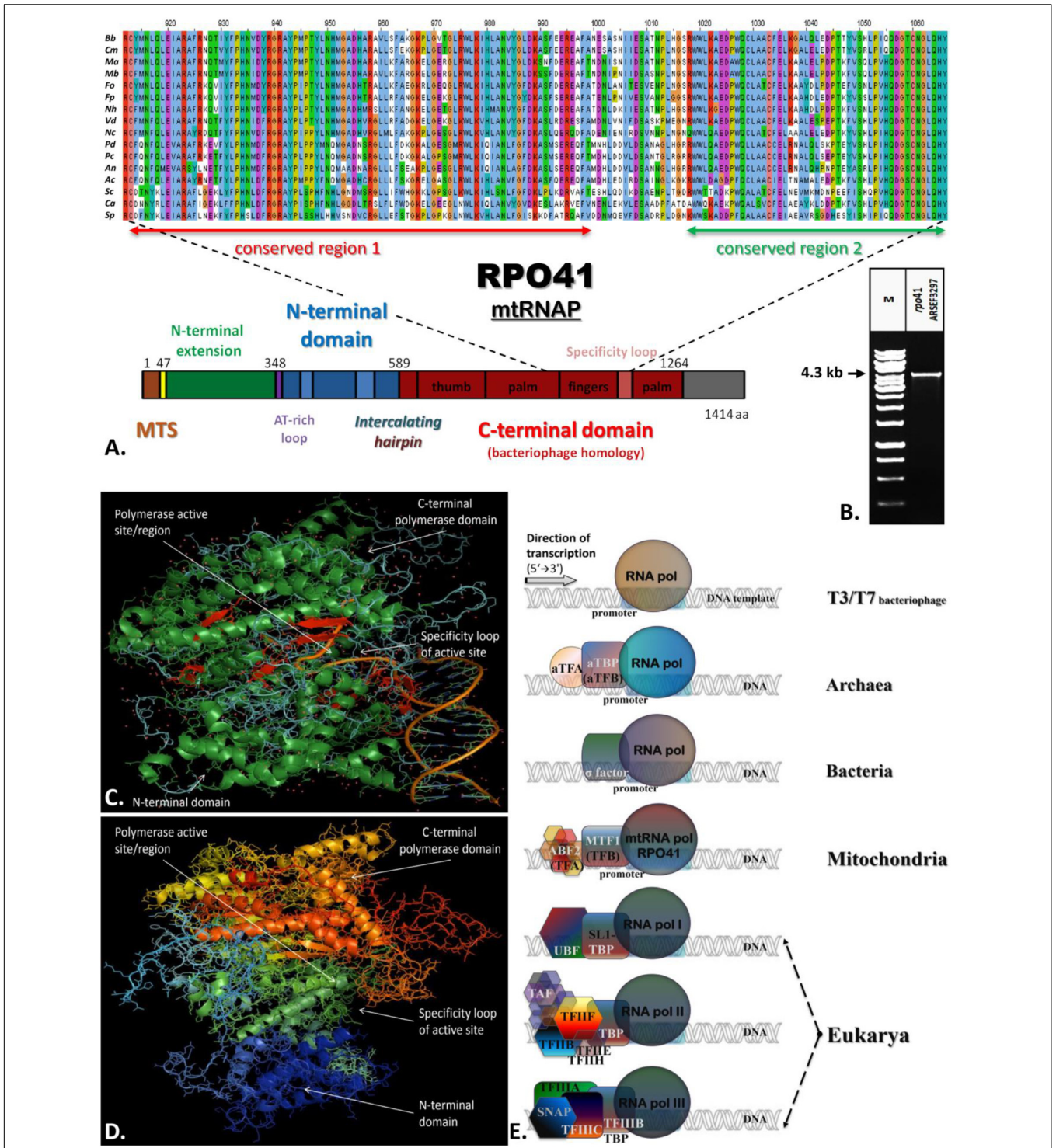


FIGURE 5 | The structure of Rpo41 from *M. brunneum* ARSEF 3297. **(A)** Multiple sequence alignment of Rpo41 from selected Ascomycetes (*Bb, Beauveria bassiana; Cm, Cordyceps militaris; Ma, Metarhizium acridium; Mb, Metarhizium brunneum; Fo, Fusarium oxysporum; Fp, Fusarium pseudograminearum; Nh, Nectria haematococca; Vd, Verticillium dahlia; Nc, Neurospora crassa; Pd, Penicillium digitatum; Pc, Penicillium chrysogenum; An, Aspergilus nidulans; Ac, Ajellomyces capsulatum; Sc, Saccharomyces cerevisiae; Ca, Candida albicans; Sp., Schizosaccharomyces pombe*), for determining conserved regions—domains, using ClustalX. **(B)** The amplified whole gene *rpo*41 of *M. brunneum* ARSEEF 3297. **(C)** Crystallographic structure of RNA polymerase is bacteriophage T7 (PDB_ 1QLN). Green represented alpha-helices, while red beta-sheets. The random coil structures are displayed in cyan. The greater part of the protein structure consists of a-helices. **(D)** Prediction of the structure of mt-RNA polymerase (Rpo41p) of *M. brunneum* ARSEF 3297. All structures were constructed using the softwares PSIPRED, HHpred and PyMol. **(E)** A comparative schematic representation of all RNA polymerases and their transcriptional factors as described for T3/T7 phages, archaea, bacteria, mitochondria, and nucleus of eukaryotes (based on information provided by Werner, 2007, 2012, Werner and Grohmann, 2011).

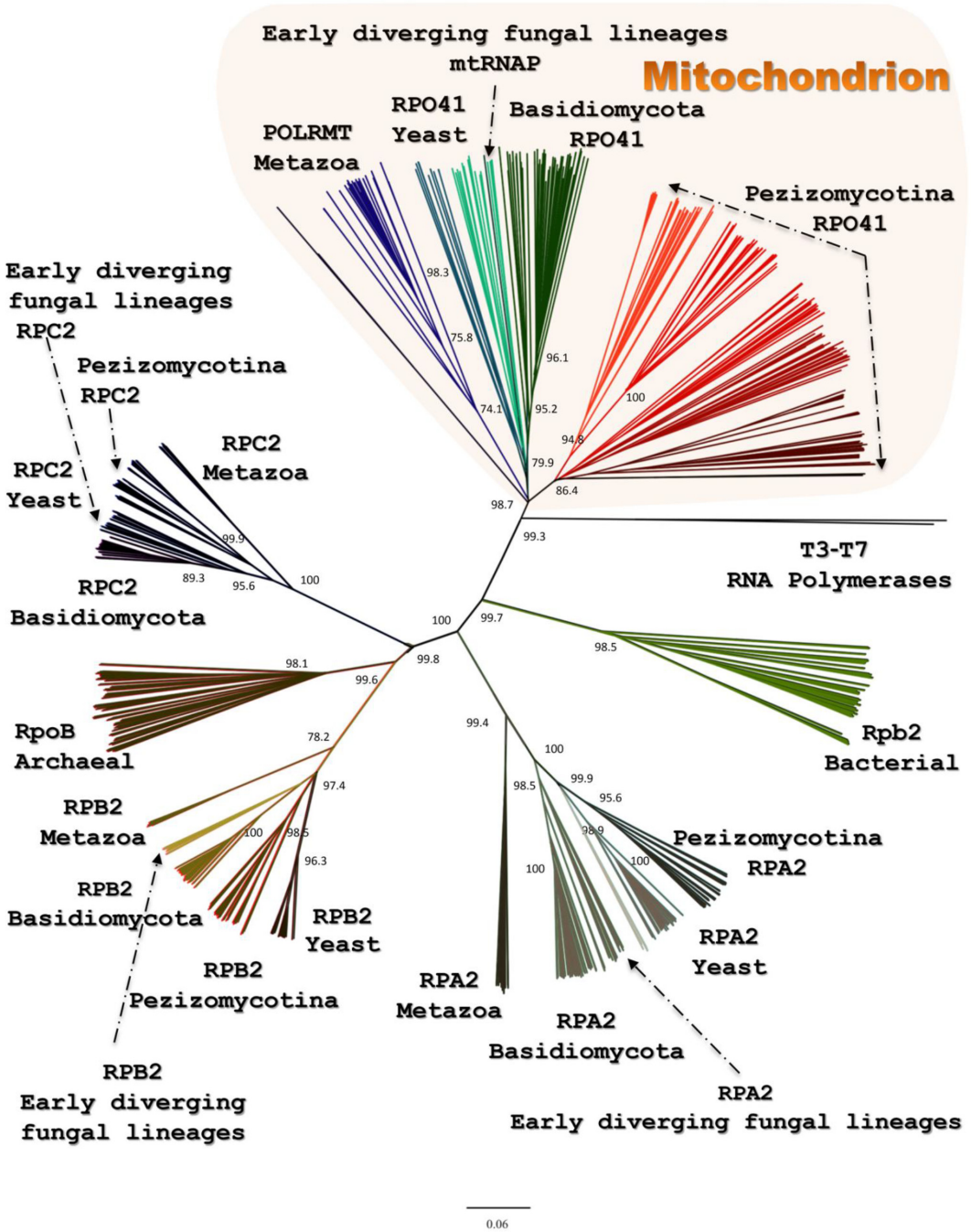


FIGURE 6 | The phylogenetic tree produced for Rpo41 and all homologs RNA polymerases from representatives of all living domains, using NJ and RaxML methods. Numbers at the nodes denote NJ-bootstrap support. RAxML-bootstrap was in all topologies above 90% and it is not shown.

with the evolutionary history of the mt transcription and its key proteins. In yeasts, many different studies have identified the mt promoter sequences, their role in the mt transcription and their distribution within the mt genome (Christianson and Rabinowitz, 1983; Biswas et al., 1985; Schinkel et al., 1986; Biswas, 1999; Schäfer et al., 2005; Kim et al., 2012; Deshpande and Patel, 2014). On the other hand, the knowledge concerning mt promoters for species belonging to Pezizomycotina is scarce, with the sole exception of *Neurospora crassa* (Burger et al., 1985; de Vries et al., 1985; Kubelik et al., 1990; Kleidon et al., 2003). Recently, a prediction of putative transcription starting sites for the mt genes of *M. anisopliae* was published (Bantihun and Kebede, 2021). In that study, the analyses were performed *in silico*, whereas in the present work, an emphasis was given

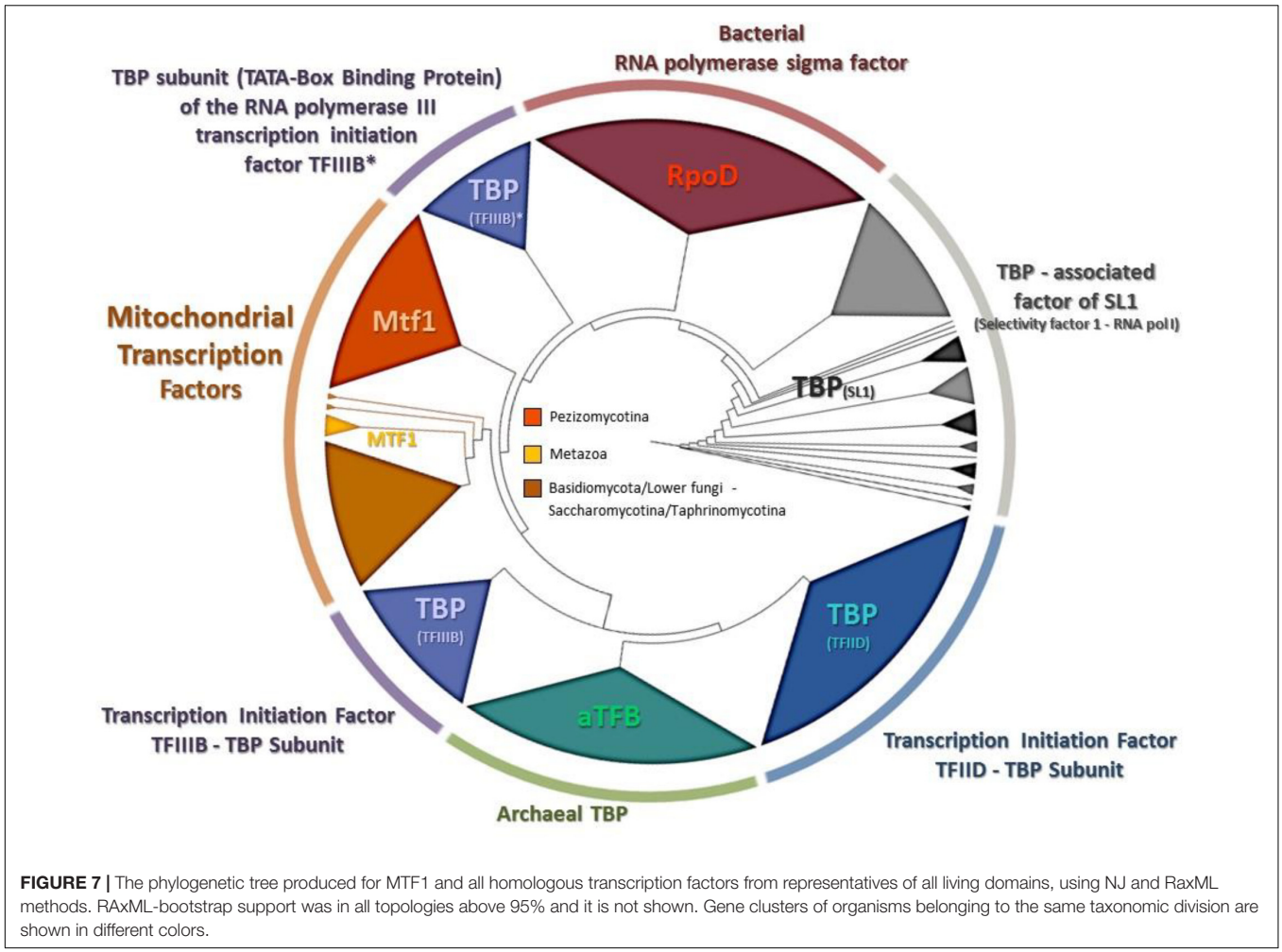


FIGURE 7 | The phylogenetic tree produced for MTF1 and all homologous transcription factors from representatives of all living domains, using NJ and RaxML methods. RAxML-bootstrap support was in all topologies above 95% and it is not shown. Gene clusters of organisms belonging to the same taxonomic division are shown in different colors.

on data based both on an *in silico* and experimental evidence, which was further combined with experimental data of the mt gene promoters of yeasts and *N. crassa*, mentioned above. Thus, the common characteristics of mt promoters are (a) their dispersal throughout the genome, (b) their participation in the transcription of polycistronic molecules which are further maturated into single transcripts and (c) their A-T rich content. In this study, both the *in silico* and the experimental data confirmed the above-mentioned characteristics of the mt promoters. In accordance to their phylogenetically closest species, i.e., *N. crassa*, the mt promoter sequence analyses of both entomopathogenic fungi of this work, presented differences to the respective sequence of yeasts. However, a few but distinctive differences of the consensus sequence from all Hypocreales were detected compared to the ones of *N. crassa*, which belongs to the Order Sordariales, especially in the 3′end of the promoter sequence. More specifically there was an underrepresentation of G in the hypocrealean mt putative consensus promoter sequence (**Figure 1**). It is worth mentioning that *in silico* prediction of the mt promoters is still difficult to be achieved, for the following reasons: (a) fungal mt genomes are AT-rich (Gray et al., 1999; Kouvelis et al., 2004), (b) gene order is variable among species of different Orders, let alone different subphyla (Kouvelis et al., 2004; Pantou et al., 2008), and (c) mt intergenic regions are variable in size and content, even between species or strains of the same genus or species, respectively (Kortsinoglou et al., 2019; Theelen et al., 2021). However, it is shown in this work, that the present abundance of sequenced mt genomes from a large number of species belonging to the same Order, like Hypocreales in this case, provides a solution for an *in silico* approach with good prediction results for determining the sequence of the mt promoter sequences. Moreover, the experimental verification of these predictions with the combined use, primarily, of RT-PCR, secondarily, of Northern Hybridization and thirdly, of 5′ end-RACE PCR, as presented in this study, showed that they may be the most suitable approaches for future similar quests. Still, many difficulties which may burden the results, have been found in this work as well as in previous studies (Burger et al., 1985; de Vries et al., 1985; Bittner-Eddy et al., 1994; Schäfer et al., 2005). In detail, mt polycistronic molecules are under the mechanism of maturation to single transcripts (Bittner-Eddy et al., 1994; Schäfer et al., 2005) and as described earlier, Northern hybridization experiments failed to provide large in size signals, as transcripts are most probably processed

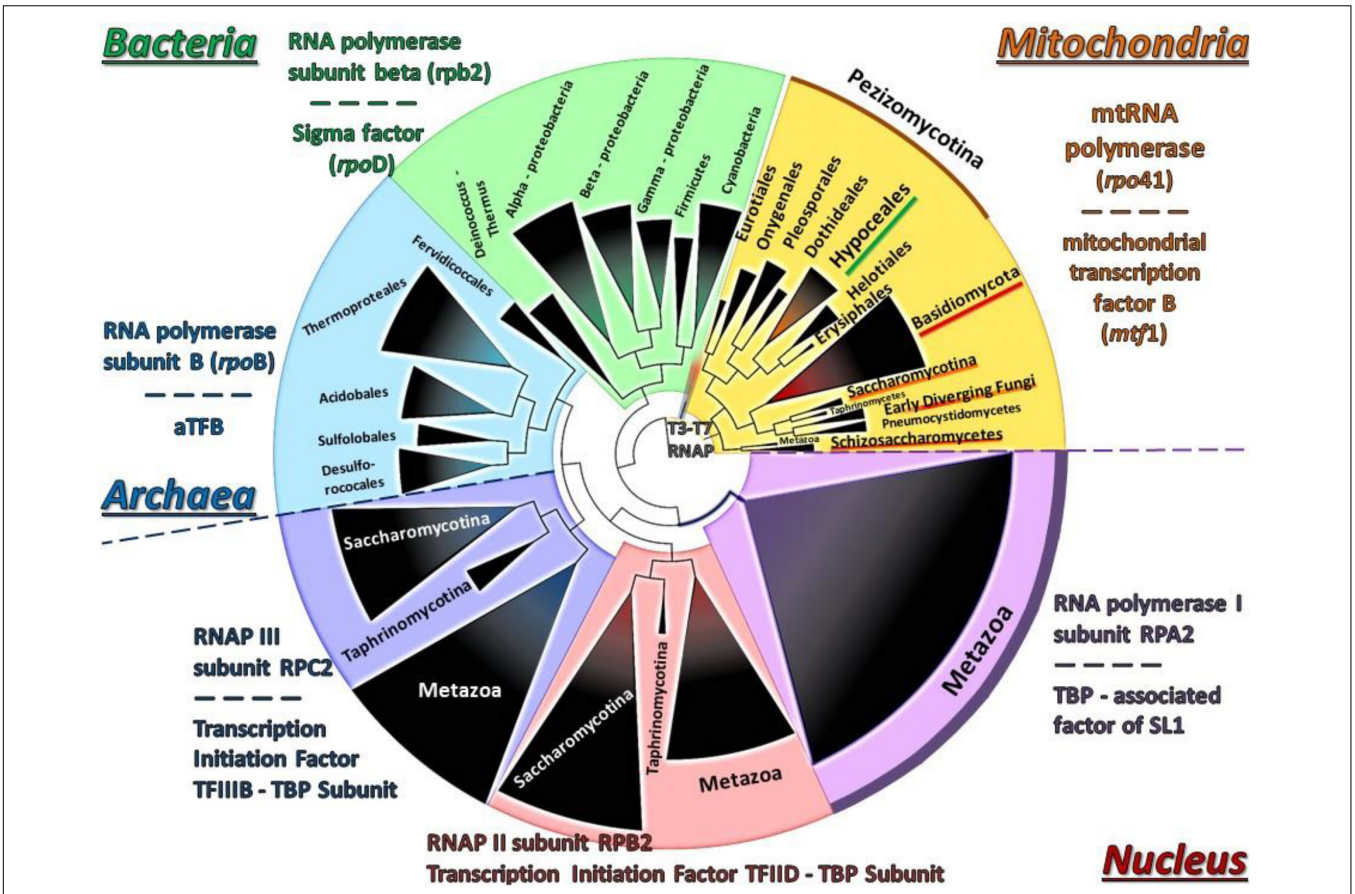


FIGURE 8 | The phylogenetic tree of the concatenated Rpo41-Mtf1 matrix using NJ and RAxML methods. RAxML-bootstrap support was in all topologies above 95% and it is not shown. Gene clusters of organisms belonging to the same taxonomic division are shown in different colors.

before transcription termination (Burger et al., 1985; de Vries et al., 1985). However, Kleidon et al. (2003) have shown that long transcripts may be retained for considerable time periods which allow their detection. RT-PCR and Northern results of this study further verify their conclusion and moreover, may provide indications about the maturation patterns of the polycistronic transcripts (**Supplementary Figure 2**). It was also shown that mt promoters of Hypocrealean fungal species with their AT-rich composition resemble the prokaryotic −10 element and the eukaryotic (nuclear) TATA box in similar evolutionary studies found in other fungi like *N. crassa* (Kleidon et al., 2003). The position of the mt promoters and the transcriptional start sites as found in this work present a resemblance to the respective elements of bacteriophage genes (Jang and Jaehning, 1991). The results of promoter sequence and the polycistronic transcripts verified the expected similarity of the mt transcription mechanism with the respective bacterial one. Therefore, these results coincide with the endosymbiotic origin of the mitochondria and their genomes (review Martin et al., 2015 and references therein).

A paradox which remains still unanswered is the origin and evolution of the mtRNA polymerase Rpo41. It is well established that the respective nuclear DNA-dependent RNA polymerases are of archaeal origin (see review Kwapisz et al. (2008) and references therein). According to the most widely accepted endosymbiotic theory of an alpha-proteobacterium as the ancestor of the mitochondrion which became a symbiont to an archaeal progenitor (reviews: Gray et al., 1999; Martin et al., 2015 and references therein), mtRNAP seems to be a descendant of a T7 phage progenitor and not of the bacterial counterpart (see review Shutt and Gray (2006) and references therein). Recently, it was shown that the C-terminal domain of Rpo41 presents homology to the catalytic domain of T7 RNAP and partially to the N-terminal domain, but it also contains a 300 amino acid region which has no similarity to the phage enzyme (Yang et al., 2015). Moreover, Rpo41 has few elements in common with the expected bacterial counterpart (Cermakian et al., 1997; Cliften et al., 2000), which are multi-subunit RNA Polymerase (msuRNAPs), closely related to their archaeal and eukaryotic homologs (Cliften et al., 2000). The structural analysis of this enzyme from the Hypocrealean entomopathogenic species of this work (**Figure 5** and **Supplementary Table 4**) and the phylogenetic tree of these enzymes (**Figure 6**) showed that mt RNA polymerases of all fungi have conserved domains with both phage and bacterial elements. However, the Rpo41 may initiate non-specific transcription, since it has retained all

promoter recognition elements found in the T7 RNAp, but in order to produce the correct polycistronic transcripts, the Mtf1 protein is needed (Yang et al., 2015). Mtf1 has all the necessary and critical amino acids found in eubacterial sigma factors, but otherwise limited homology as whole proteins (Cermakian et al., 1997). Therefore, in this study, an effort was made to concatenate both proteins in one matrix (**Figure 8**), including sequences from representative RNAPs from all domains of life, i.e., phages, archaea, bacteria, eukaryotes (with sequences of both nuclear and mitochondrial origin), in order to define their phylogenetic signal, as it is well established that single genie phylogenies do not always represent the phylogenetic relationships or the evolution of whole organisms or cluster of genes involved in one process (e.g., Pantou et al., 2008; Korovesi et al., 2018). However, single protein phylogenies of this work were in agreement to the previously mentioned studies. Although the tree of the concatenated dataset must be treated cautiously, due to the lack of congruency support, it showed that the fungal holoenzyme of mtRpo41-MTF1 remains basal to the tree and close to the single unit T7/T3 RNA polymerase which does not need any factor for the transcription of its genome (**Figure 8**). Moreover, based on the review of de Juan et al. (2013), three different methodologies, i.e., the evolutionary trace (SDP method), STRING (method of phylogenetic profiles), and the mirror tree (an inter-protein co-evolution method), were employed and verified that these proteins remained conserved and co-evolved through evolution, despite their differential origin.

A possible mechanism for the acquisition and usage of the Rpo41, the homolog ssuRNAP of T3/T7 phage, instead of the original multi-subunit bacteria-like RNA polymerase, is the inheritance of a bacteriophage T3/T7-like RNA polymerase at the proto-mitochondrial ancestor in a later (secondarily) stage (Burger and Lang, 2003). While this hypothesis cannot be ruled out, in this work phylogenetic analyses, as well as the comparative enzyme structure analyses of these critical to mt transcription proteins, i.e., Rpo41 and MTF1, showed the following: (a) the phylogenetic positioning of the main transcriptional factor MTF1 confirmed the above results as it was shown that MTF1, due to its sequence alignment, remains sister clade to the transcription factors of bacteria which recognize −10 sequences and the eukaryotic factors which need TATA boxes in order to proceed to transcription of the genes (**Figure 7**), (b) the resemblance of mtRPO41 with the T7/T3 respective enzyme, based on this analyses of an immense number of sequences compared to the original studies was verified (**Figures 5**, **6** and **Supplementary Tables 4**, **5**), and (c) the presence of both enzymes in all genomes analyzed, a most probable result of vertical inheritance from the early evolution of the eukaryote. Thus, we may support that during the endosymbiotic incident that led to the genesis of the mitochondrion, the progenitor of the mitochondria might have been an ancestral bacterium which was infected with a phage-like progenitor. This hypothesis has been also considered as the most plausible scenario for explaining the role of this phage-like polymerase in mitochondria (Filée and Forterre, 2005), and may be considered as a variation of the "Viral Eukaryogenesis" (VE) hypothesis, proposed by Bell (2001, 2006) and Forterre (2006). However, VE hypothesis proposes that the first eukaryotic cell was a "simultaneous" consortium of three organisms i.e., an archaeon, an a-proteobacterium and a viral ancestor of the nucleus (Bell, 2009). Additionally, in its current version, the unique features of the nucleus, like the uncoupling of transcription from translation, are of viral origin (Bell, 2020). Even when the objections (Koonin and Yutin, 2018; Krupovic et al., 2019; Slijepcevic, 2021) for the acceptance of this hypothesis are considered, our hypothesis of a mitochondrion originating from an a-proteobacterial ancestor infected by a phage-like progenitor cannot be overruled and seems an alternative new variation of the prophage infected alpha-proteobacerium which acted as the mitochondrial progenitor (Filée and Forterre, 2005). Additionally, the results of this work based on the parallel genetic analyses of the genomes' transcriptional elements and the phylogenetic analyses of transcription's key enzymes (as presented above) are in accordance to results of previous studies (Masters et al., 1987; Tracy and Stern, 1995; Cliften et al., 2000), thus supporting further the hypothesis proposed by Filée and Forterre (2005). This work further improves this latter hypothesis by pinpointing the co-operation of phage (RNA polymerase) and a bacterial (sigma-like Mtf1) originated components for the correct transcription of mt genes. The alternative possible hypothesis that the acquisition of an ancestral polymerase related to phages is a result of Horizontal Gene Transfer is weak based on the data of this work, since there are indications mentioned above which show that the mt polymerase and transcription factor were included in the mt gene transcription from the beginning and coevolved from the early start of the mitochondrion's formation.

CONCLUSION

In conclusion, the mt transcription of the fungal entomopathogenic species *B. bassiana* and *M. brunneum* was analyzed in this study, concerning the promoters involved, the polycistronic transcripts which are transcribed and their possible maturation, the structure and phylogeny of the key enzymes of this transcription, i.e., Rpo41 and MTF1. It was shown that there is a certain diversity of the promoter sites and of the content of the polycistronic transcripts among the different species, especially between yeasts and Pezizomycotina. However, a consensus is obvious to the key enzymes and the mechanism of transcription. If the phylogenetic analyses are expanded to all RNA polymerases and their transcriptional factors from representatives from all domains of life, then this comparative analysis provides insights to the evolution of the mitochondrion and the mechanism of transcription of its genome. In detail, it suggests that the mechanism of fungal mt transcription is similar to the bacterial respective process with the inclusion of a single transcription factor. The RNA polymerase is of phage origin and, as suggested previously by other studies, might have been acquired in a very early stage of the mitochondrion's formation. That is, the bacterial endosymbiont was infected with a phage and for reasons which need further examination, this polymerase of the phage ancestor

remained in the mitochondrion progenitor as the functional one. This polymerase, however, coevolved further with the existing primitive sigma factor which evolved to the MTF1 factor.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

SV and VK conceived, designed the study, and analyzed the data. SV conducted the experiments and wrote the manuscript. VK revised the manuscript. Both authors read and approved the final manuscript.

FUNDING

This research work has been financed by the Hellenic Foundation for Research and Innovation (HFRI) under the HFRI Ph.D. Fellowship grant (Fellowship No. 16177) and the Special Account for Research Grand (SARG) of the NKUA (Grant No. 11239).

ACKNOWLEDGMENTS

Metarhizium brunneum ARSEF 3297 and *Beauveria bassiana* ATHUM 4946 were kindly provided by T. M. Butt (Swansea University, College of Science, Department of Biosciences, Wales, United Kingdom) and Z. Gonou—Zagou (University of Athens, Faculty of Biology, Dept. of Ecology and Systematics, Athens, Greece), respectively. SV and VK thank HFRI for the funding of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.821638/full#supplementary-material

Supplementary Figure 1 | The map of the mitochondrial genome of *Metarhizium brunneum* ARSEF 3297. Arrows indicate the direction of gene transcription. The inner circles show the GC content. All genes identified are indicated in italics.

Supplementary Figure 2 | Northern blots of mitochondrial transcriptome and gene synteny of mt genome for *B. bassiana*. RNA gel—blot hybridizations with comprehensive sets of labeled oligonucleotides covering the sense DNA strand in *B. bassiana* ATHUM 4946: (I) *rnl*, (II) *nad*3, (III) *cob*, (IV) *cox*1, (V) *nad*4, (VI) *cox*3. Inferred processing pathway for transcripts of the polycistronic transcription units in mitochondria from *B. bassiana*. The above pathways result from an analysis of the transcripts that behave as cohorts (six polycistronic transcription units) and are detected by related hybridization probes, in RNA samples from spores at different stages of germination. The triangle symbol denotes a nuclease processing step that occurs for the gene-specific post-transcriptional mechanism in the expression of protein-coding genes in Hypocrealean mitochondria. Transcripts, exons and introns are not drawn to scale. The direction of transcription is from left to right.

Supplementary Figure 3 | The phylogenetic tree produced for Rpo41 and all homologs RNA polymerases from representatives of all living domains, using NJ and RaxML methods. Numbers at the nodes denote NJ-bootstrap support. RAxML-bootstrap was in all topologies above 90% and it is not shown. OTUs are presented as acronyms whose full details are provided in **Supplementary Table 3**.

Supplementary Figure 4 | The phylogenetic tree produced for MTF1 and all homologous transcription factors from representatives of all living domains, using NJ and RaxML methods. RAxML-bootstrap support was in all topologies above 95% and it is not shown. Gene clusters of organisms belonging to the same taxonomic division are shown in different colors. OTUs are presented as acronyms whose full details are provided in **Supplementary Table 3**.

Supplementary Figure 5 | The phylogenetic tree of the concatenated Rpo41-Mtf1 matrix using NJ and RAxML methods. RAxML-bootstrap support was in all topologies above 95% and it is not shown. Gene clusters of organisms belonging to the same taxonomic division are shown in different colors. OTUs are presented as acronyms whose full details are provided in **Supplementary Table 3**.

Supplementary Figure 6 | The comparison of the two single phylogenetic trees of Rpo41 and Mtf1 as produced by MirrorTree and showed by Phylo.io. The species included are shown in numbers (according to the enumeration of species in PBD databank) as follows: *Ashbya gossypii* 284811, *Candida glabrata* 284593, *Kluyveromyces lactis* 284590, *Komagataella phaffii* 644223, *Lachancea thermotolerans* 559295, *Meyerozyma guillermondii* 294746, *S. cerevisiae* 285006, 307796, 559292, 574961, 643680, and *Zygosaccharomyces rouxii* 559307.

Supplementary Table 1 | The annotated complete mt-genome of *M. brunneum* ARSEF 3297.

Supplementary Table 2 | Annotation attributes of the complete mt genomes of *M. brunneum* strains ARSEF 3297 and ARSEF 4556. Cells highlighted in yellow indicate intronic ORFs. Genes encoding rRNAs, and subunits of ATP synthase, Apocytochrome b, Cytochrome oxidase c and NADH dehydrogenase are shown as highlighted cells in blue, orange, light gray, purple and light green, respectively.

Supplementary Table 3 | The representative species from all domains of life for which the RNA polymerases and their Transcription factors were used in the phylogenetic analyses. Their taxonomic status of these species, the GenBank Accession Numbers of their sequences are provided.

Supplementary Table 4 | Conserved important residues of Rpo41 and Mtf1 proteins as presented by ET1052 viewer and the analyses of De Wijngaert et al., 2021.

Supplementary Table 5 | Predicted Complex Interaction of Rpo41 and MTF1 as found by STRING method.

REFERENCES

Aguileta, G., De Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., et al. (2014). High variability of mitochondrial gene order among fungi. *Genome Biol. Evol.* 6, 451–465. doi: 10.1093/gbe/evu028

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Bantihun, G., and Kebede, M. (2021). *In silico* analysis of promoter region and regulatory elements of mitogenome co-expressed *trn* gene clusters encoding for bio-pesticide in entomopathogenic fungus, *Metarhizium anisopliae*: strain ME1. *J. Genet. Eng. Biotechnol.* 19:94. doi: 10.1186/s43141-021-00191-6

Basu, U., Mishra, N., Farooqui, M., Shen, J., Johnson, L. C., and Patel, S. S. (2020). Auto-inhibitory regulation of DNA binding by the C-terminal tails of the mitochondrial transcription factors Mtf1 and TFB2M. *bioRxiv* [Preprint]. doi: 10.1101/2020.03.06.980961

Bell, P. (2020). Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res.* 289:198168. doi: 10.1016/j.virusres.2020.198168

Bell, P. J. (2001). Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* 53, 251–256. doi: 10.1007/s002390010215

Bell, P. J. (2006). Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. *J. Theor. Biol.* 243, 54–63. doi: 10.1016/j.jtbi.2006.05.015

Bell, P. J. (2009). The viral eukaryogenesis hypothesis: a key role for viruses in the emergence of eukaryotes from a prokaryotic world environment. *Ann. N. Y. Acad. Sci.* 1178, 91–105. doi: 10.1111/j.1749-6632.2009.04994.x

Biswas, T. K. (1999). Nucleotide sequences surrounding the nonanucleotide promoter motif influence the activity of yeast mitochondrial promoter. *Biochemistry* 38, 9693–9703. doi: 10.1021/bi982804l

Biswas, T. K., Edwards, J. C., Rabinowitz, M., and Getz, G. S. (1985). Characterization of a yeast mitochondrial promoter by deletion mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* 82, 1954–1958. doi: 10.1073/pnas.82.7.1954

Bittner-Eddy, P., Monroy, A. F., and Brambl, R. (1994). Expression of mitochondrial genes in the germinating conidia of *Neurospora crassa*. *J. Mol. Biol.* 235, 881–897. doi: 10.1006/jmbi.1994.1046

Bonawitz, N. D., Clayton, D. A., and Shadel, G. S. (2006). Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Mol. Cell* 24, 813–825. doi: 10.1016/j.molcel.2006.11.024

Borukhov, S., and Severinov, K. (2002). Role of the RNA polymerase sigma subunit in transcription initiation. *Res. Microbiol.* 153, 557–562. doi: 10.1016/S0923-2508(02)01368-2

Breitenberger, C. A., Browning, K. S., Alzner-DeWeerd, B., and RajBhandary, U. L. (1985). RNA processing in *Neurospora crassa* mitochondria: use of transfer RNA sequences as signals. *EMBO J.* 4, 185–195. doi: 10.1002/j.1460-2075.1985.tb02335.x

Buchan, D. W., and Jones, D. T. (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* 47, W402–W407. doi: 10.1093/nar/gkz297

Bullerwell, C. E., Burger, G., and Lang, B. F. (2000). A novel motif for identifying *rps*3 homologs in fungal mitochondrial genomes. *Trends Biochem. Sci.* 25, 363–365. doi: 10.1016/S0968-0004(00)01612-1

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716. doi: 10.1016/j.tig.2003.10.012

Burger, G., Helmer Citterich, M., Nelson, M. A., Werner, S., and Macino, G. (1985). RNA processing in *Neurospora crassa* mitochondria: transfer RNAs punctuate a large precursor transcript. *EMBO J.* 4, 197–204. doi: 10.1002/j.1460-2075.1985.tb02336.x

Burger, G., and Lang, B. F. (2003). Parallels in genome evolution in mitochondria and bacterial symbionts. *IUBMB Life* 55, 205–212. doi: 10.1080/1521654031000137380

Butow, R. A., Zhu, H., Perlman, P., and Conrad-Webb, H. (1989). The role of a conserved dodecamer sequence in yeast mitochondrial gene expression. *Genome* 31, 757–760. doi: 10.1139/g89-134

Butt, T. M., Coates, C. J., Dubovskiy, I. M., and Ratcliffe, N. A. (2016). Entomopathogenic fungi: new insights into host–pathogen interactions. *Adv. Genet.* 94, 307–364. doi: 10.1016/bs.adgen.2016.01.006

Butt, T. M., Jackson, C., and Magan, N. (eds) (2001). *Fungi as Biocontrol Agents: Progress Problems and Potential*. Wallingford: CABI. doi: 10.1079/9780851993560.0000

Cech, T. R. (1990). Self-splicing of group I introns. *Annu. Rev. Biochem.* 59, 543–568. doi: 10.1146/annurev.bi.59.070190.002551

Cermakian, N., Ikeda, T. M., Miramontes, P., Lang, B. F., Gray, M. W., and Cedergren, R. (1997). On the evolution of the single-subunit RNA polymerases. *J. Mol. Evol.* 45, 671–681. doi: 10.1007/pl00006271

Chan, P. P., and Lowe, T. M. (2019). "tRNAscan-SE: searching for tRNA genes in genomic sequences," in *Gene Prediction. Methods in Molecular Biology*, Vol. 1962, ed. M. Kollmar (New York, NY: Humana Press), 1–14. doi: 10.1007/978-1-4939-9173-0_1

Christianson, T., and Rabinowitz, M. (1983). Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by *in vitro* capping with guanylyltransferase. *J. Biol. Chem.* 258, 14025–14033. doi: 10.1016/S0021-9258(17)44019-1

Christianson, T. W., and Clayton, D. A. (1988). A tridecamer DNA sequence supports human mitochondrial RNA 3'-end formation *in vitro*. *Mol. Cell. Biol.* 8, 4502–4509. doi: 10.1128/mcb.8.10.4502-4509.1988

Christinaki, A. C., Kanellopoulos, S. G., Kortsinoglou, A. M., Theelen, B., Boekhout, T., and Kouvelis, V. N. (2021). Mitogenomics and mitochondrial gene phylogeny decipher the evolution of *Saccharomycotina* yeasts. *bioRxiv* [Preprint]. doi: 10.1101/2021.06.11.448017

Clarkson, J. M., and Charnley, A. K. (1996). New insights into the mechanisms of fungal pathogenesis in insects. *Trends Microbiol.* 4, 197–203. doi: 10.1016/0966-842X(96)10022-6

Clayton, D. A. (1991). Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell Biol.* 7, 453–478. doi: 10.1146/annurev.cb.07.110191.002321

Cliften, P. F., Jang, S. H., and Jaehning, J. A. (2000). Identifying a core RNA polymerase surface critical for interactions with a sigma-like specificity factor. *Mol. Cell. Biol.* 20, 7013–7023. doi: 10.1128/MCB.20.18.7013-7023.2000

Costanzo, M. C., and Fox, T. D. (1990). Control of mitochondrial gene expression in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* 24, 91–113. doi: 10.1146/annurev.ge.24.120190.000515

Cotney, J., and Shadel, G. S. (2006). Evidence for an early gene duplication event in the evolution of the mitochondrial transcription factor B family and maintenance of rRNA methyltransferase activity in human mtTFB1 and mtTFB2. *J. Mol. Evol.* 63, 707–717. doi: 10.1007/s00239-006-0075-1

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

de Juan, D., Pazos, F., and Valencia, A. (2013). Emerging methods in protein co-evolution. *Nat. Rev. Genet.* 14, 249–261. doi: 10.1038/nrg3414

de Vries, H., Haima, P., Brinker, M., and de Jonge, J. C. (1985). The *Neurospora* mitochondrial genome: the region coding for the polycistronic cytochrome oxidase subunit I transcript is preceded by a transfer RNA gene. *FEBS Lett.* 179, 337–342. doi: 10.1016/0014-5793(85)80547-0

De Wijngaert, B., Sultana, S., Singh, A., Dharia, C., Vanbuel, H., Shen, J., et al. (2021). Cryo-EM structures reveal transcription initiation steps by yeast mitochondrial RNA polymerase. *Mol. Cell* 81, 268–280. doi: 10.1016/j.molcel.2020.11.016

Deshpande, A. P., and Patel, S. S. (2012). Mechanism of transcription initiation by the yeast mitochondrial RNA polymerase. *Biochim. Biophys. Acta Gene Regul. Mech.* 1819, 930–938. doi: 10.1016/j.bbagrm.2012.02.003

Deshpande, A. P., and Patel, S. S. (2014). Interactions of the yeast mitochondrial RNA polymerase with the+ 1 and+ 2 promoter bases dictate transcription initiation efficiency. *Nucleic Acids Res.* 42, 11721–11732. doi: 10.1093/nar/gku868

Dieckmann, C. L., and Staples, R. R. (1994). Regulation of mitochondrial gene expression in *Saccharomyces cerevisiae*. *Int. Rev. Cytol.* 152, 145–181. doi: 10.1016/s0074-7696(08)62556-5

Filée, J., and Forterre, P. (2005). Viral proteins functioning in organelles: a cryptic origin? *Trends Microbiol.* 13, 510–513. doi: 10.1016/j.tim.2005.08.012

Fonseca, P. L., Badotti, F., De-Paula, R. B., Araújo, D. S., Bortolini, D. E., Del-Bem, L. E., et al. (2020). Exploring the relationship among divergence time and coding and non-coding elements in the shaping of fungal mitochondrial genomes. *Front. Microbiol.* 11:765. doi: 10.3389/fmicb.2020.00765

Forterre, P. (2006). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3669–3674. doi: 10.1073/pnas.0510333103

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

Gagliardi, D., Stepien, P. P., Temperley, R. J., Lightowlers, R. N., and Chrzanowska-Lightowlers, Z. M. (2004). Messenger RNA stability in mitochondria: different means to an end. *Trends Genet.* 20, 260–267. doi: 10.1016/j.tig.2004.04.006

Gao, Q., Jin, K., Ying, S. H., Zhang, Y., Xiao, G., Shang, Y., et al. (2011). Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet.* 7:e1001264. doi: 10.1371/journal.pgen.1001264

Ghikas, D. V., Kouvelis, V. N., and Typas, M. A. (2006). The complete mitochondrial genome of the entomopathogenic fungus *Metarhizium anisopliae* var. *anisopliae*: gene order and *trn* gene clusters reveal a common

evolutionary course for all *Sordariomycetes*. *Arch. Microbiol.* 185, 393–401. doi: 10.1007/s00203-006-0104-x

Ghikas, D. V., Kouvelis, V. N., and Typas, M. A. (2010). Phylogenetic and biogeographic implications inferred by mitochondrial intergenic region analyses and ITS1-5.8S-ITS2 of the entomopathogenic fungi *Beauveria bassiana* and *B. brongniartii*. *BMC Microbiol.* 10:174. doi: 10.1186/1471-2180-10-174

Glare, T., Campbell, M., Biggs, P., Winter, D., Durrant, A., McKinnon, A., et al. (2020). Mitochondrial evolution in the entomopathogenic fungal genus *Beauveria*. *Arch. Insect Biochem. Physiol.* 105:e21754. doi: 10.1002/arch.21754

Gray, M. W., Burger, G., and Lang, B. F. (1999). Mitochondrial evolution. *Science* 283, 1476–1481. doi: 10.1126/science.283.5407.1476

Grivell, L. A. (1995). Nucleo-mitochondrial interactions in mitochondrial gene expression. *Crit. Rev. Biochem. Mol. Biol.* 30, 121–164. doi: 10.3109/10409239509085141

Hausner, G. (2003). Fungal mitochondrial genomes. *Fungal Genom.* 3:101.

Hofmann, T. J., Min, J., and Zassenhaus, H. P. (1993). Formation of the 3' end of yeast mitochondrial mRNAs occurs by site-specific cleavage two bases downstream of a conserved dodecamer sequence. *Yeast* 9, 1319–1330. doi: 10.1002/yea.320091205

Hu, X., Xiao, G., Zheng, P., Shang, Y., Su, Y., Zhang, X., et al. (2014). Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16796–16801. doi: 10.1073/pnas.1412662111

Jang, S. H., and Jaehning, J. A. (1991). The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *J. Biol. Chem.* 266, 22671–22677. doi: 10.1016/S0021-9258(18)54622-6

Jiang, H., Sun, W., Wang, Z., Zhang, J., Chen, D., and Murchie, A. I. (2011). Identification and characterization of the mitochondrial RNA polymerase and transcription factor in the fission yeast *Schizosaccharomyces pombe*. *Nucleic Acids Res.* 39, 5119–5130. doi: 10.1093/nar/gkr103

Kennell, J. C., and Lambowitz, A. M. (1989). Development of an *in vitro* transcription system for *Neurospora crassa* mitochondrial DNA and identification of transcription initiation sites. *Mol. Cell. Biol.* 9, 3603–3613. doi: 10.1128/mcb.9.9.3603-3613.1989

Kim, H., Tang, G. Q., Patel, S. S., and Ha, T. (2012). Opening–closing dynamics of the mitochondrial transcription pre-initiation complex. *Nucleic Acids Res.* 40, 371–380. doi: 10.1093/nar/gkr736

Kleidon, J., Plesofsky, N., and Brambl, R. (2003). Transcripts and transcript-binding proteins in mitochondria of *Neurospora crassa*. *Mitochondrion* 2, 345–360. doi: 10.1016/S15677249(03)00002-3

Kolondra, A., Labedzka-Dmoch, K., Wenda, J. M., Drzewicka, K., and Golik, P. (2015). The transcriptome of *Candida albicans* mitochondria and the evolution of organellar transcription units in yeasts. *BMC Genom.* 16:827. doi: 10.1186/s12864-015-2078-z

Koonin, E. V., and Yutin, N. (2018). Multiple evolutionary origins of giant viruses. *F1000Research* 7:F1000 Faculty Rev-1840. doi: 10.12688/f1000research.16248.1

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

Kortsinoglou, A. M., Korovesi, A. G., Theelen, B., Hagen, F., Boekhout, T., and Kouvelis, V. N. (2019). The mitochondrial intergenic regions *nad*1-*cob* and *cob-rps*3 as molecular identification tools for pathogenic members of the genus *Cryptococcus*. *FEMS Yeast Res.* 19:foz077. doi: 10.1093/femsyr/foz077

Kortsinoglou, A. M., Saud, Z., Eastwood, D. C., Butt, T. M., and Kouvelis, V. N. (2020). The mitochondrial genome contribution to the phylogeny and identification of *Metarhizium* species and strains. *Fungal Biol.* 124, 845–853. doi: 10.1016/j.funbio.2020.06.003

Kouvelis, V. N., Ghikas, D. V., Edgington, S., Typas, M. A., and Moore, D. (2008a). Molecular characterisation of isolates of *Beauveria bassiana* obtained from overwintering and summer populations of Sunn Pest (*Eurygaster integriceps*). *Lett. Appl. Microbiol.* 46, 414–420. doi: 10.1111/j.1472-765X.2008.02331.x

Kouvelis, V. N., Sialakouma, A., and Typas, M. A. (2008b). Mitochondrial gene sequences alone or combined with ITS region sequences provide firm molecular criteria for the classification of *Lecanicillium* species. *Mycol. Res.* 112, 829–844. doi: 10.1016/j.mycres.2008.01.016

Kouvelis, V. N., Ghikas, D. V., and Typas, M. A. (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, 930–940. doi: 10.1016/j.fgb.2004.07.003

Kouvelis, V. N., Wang, C., Skrobek, A., Pappas, K. M., Typas, M. A., and Butt, T. M. (2011). Assessing the cytotoxic and mutagenic effects of secondary metabolites produced by several fungal biological control agents with the Ames assay and the VITOTOX® test. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 722, 1–6. doi: 10.1016/j.mrgentox.2011.01.004

Krupovic, M., Dolja, V. V., and Koonin, E. V. (2019). Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* 17, 449–458. doi: 10.1038/s41579-019-0205-6

Kubelik, A. R., Kennell, J. C., Akins, R. A., and Lambowitz, A. M. (1990). Identification of *Neurospora* mitochondrial promoters and analysis of synthesis of the mitochondrial small rRNA in wild-type and the promoter mutant [poky]. *J. Biol. Chem.* 265, 4515–4526. doi: 10.1016/S0021-9258(19)39593-6

Kwapisz, M., Beckouët, F., and Thuriaux, P. (2008). Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet.* 24, 211–215. doi: 10.1016/j.tig.2008.02.002

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

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lipinski, K. A., 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 Bioenerget.* 1797, 1086–1098. doi: 10.1016/j.bbabio.2009.12.019

Lovett, B., and Leger, R. S. (2016). *Genetics and Molecular Biology of Entomopathogenic Fungi*. Cambridge, MA: Academic Press.

Martin, W. F., Garg, S., and Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140330. doi: 10.1098/rstb.2014.0330

Masters, B. S., Stohl, L. L., and Clayton, D. A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* 51, 89–99. doi: 10.1016/0092-8674(87)90013-4

McAllister, W. T., and Raskin, C. A. (1993). The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol. Microbiol.* 10, 1–6. doi: 10.1111/j.1365-2958.1993.tb00897.x

Megarioti, A. H., and Kouvelis, V. N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.* 12, 1337–1354. doi: 10.1093/gbe/evaa126

Michel, F., and Ferat, J. L. (1995). Structure and activities of group II introns. *Annu. Rev. Biochem.* 64, 435–461. doi: 10.1146/annurev.bi.64.070195.002251

Morgan, D. H., Kristensen, D. M., Mittleman, D., and Lichtarge, O. (2006). ET viewer: an application for predicting and visualizing functional sites in protein structures. *Bioinformatics* 22, 2049–2050. doi: 10.1093/bioinformatics/btl285

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

Ochoa, D., and Pazos, F. (2010). Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics* 26, 1370–1371. doi: 10.1093/bioinformatics/btq137

Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470–474. doi: 10.1038/290470a0

Osinga, K. A., De Vries, E., Van der Horst, G., and Tabak, H. F. (1984). Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. *EMBO J.* 3, 829–834. doi: 10.1002/j.1460-2075.1984.tb01892.x

Osinga, K. A., Haan, M. D., Christianson, T., and Tabak, H. F. (1982). A nonanucleotide sequence involved in promotion of ribosomal RNA synthesis and RNA priming of DNA replication on yeast mitochondria. *Nucleic Acids Res.* 10, 7993–8006. doi: 10.1093/nar/10.24.7993

Pantou, M. P., Kouvelis, V. N., and Typas, M. A. (2008). The complete mitochondrial genome of *Fusarium oxysporum*: insights into fungal mitochondrial evolution. *Gene* 419, 7–15. doi: 10.1016/j.gene.2008.04.009

Paquin, B., Laforest, M. J., Forget, L., Roewer, I., Wang, Z., Longcore, J., et al. (1997). The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr. Genet.* 31, 380–395. doi: 10.1007/s002940050220

Paratkar, S., and Patel, S. S. (2010). Mitochondrial transcription factor Mtf1 traps the unwound non-template strand to facilitate open complex formation 2. *J. Biol. Chem.* 285, 3949–3956. doi: 10.1074/jbc.M109.050732

Pel, H. J., and Grivell, L. A. (1993). The biology of yeast mitochondrial introns. *Mol. Biol. Rep.* 18, 1–13. doi: 10.1007/BF01006890

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

Ramachandran, A., Nandakumar, D., Deshpande, A. P., Lucas, T. P., Ramanagouda, R., Tang, G. Q., et al. (2016). The yeast mitochondrial RNA polymerase and transcription factor complex catalyzes efficient priming of DNA synthesis on single-stranded DNA. *J. Biol. Chem.* 291, 16828–16839. doi: 10.1074/jbc.M116.740282

Robinson, O., Dylus, D., and Dessimoz, C. (2016). Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol. Biol. Evol.* 33, 2163–2166. doi: 10.1093/molbev/msw080

Rodeheffer, M., Boone, B., Bryan, A., and Shadel, G. (2001). Nam1p, a protein involved in RNA processing and translation, is coupled to transcription through an interaction with yeast mitochondrial RNA polymerase. *J. Biol. Chem.* 276, 8616–8622. doi: 10.1074/jbc.M009901200

Rorbach, J., and Minczuk, M. (2012). The post-transcriptional life of mammalian mitochondrial RNA. *Biochem. J.* 444, 357–373. doi: 10.1042/BJ20112208

Rorbach, J., Bobrowicz, A., Pearce, S., and Minczuk, M. (2014). "Polyadenylation in bacteria and organelles," in *Polyadenylation. Methods in Molecular Biology (Methods and Protocols)*, Vol. 1125, eds J. Rorbach and A. Bobrowicz (Totowa, NJ: Humana Press). doi: 10.1007/978-1-62703-971-0_18

Saud, Z., Kortsinoglou, A. M., Kouvelis, V. N., and Butt, T. M. (2021). Telomere length de novo assembly of all 7 chromosomes and mitogenome sequencing of the model entomopathogenic fungus, *Metarhizium brunneum*, by means of a novel assembly pipeline. *BMC Genomics* 22:87. doi: 10.1186/s12864-021-07390-y

Savkina, M., Temiakov, D., McAllister, W. T., and Anikin, M. (2010). Multiple functions of yeast mitochondrial transcription factor Mtf1p during initiation 2. *J. Biol. Chem.* 285, 3957–3964. doi: 10.1074/jbc.M109.051003

Schäfer, B. (2005). RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe*. *Gene* 354, 80–85. doi: 10.1016/j.gene.2005.03.032

Schäfer, B., Hansen, M., and Lang, B. F. (2005). Transcription and RNA-processing in fission yeast mitochondria. *RNA* 11, 785–795. doi: 10.1261/rna.7252205

Schinkel, A. H., Groot Koerkamp, M. J., Van der Horst, G. T., Touw, E. P., Osinga, K. A., Van der Bliek, A. M., et al. (1986). Characterization of the promoter of the large ribosomal RNA gene in yeast mitochondria and separation of mitochondrial RNA polymerase into two different functional components. *EMBO J.* 5, 1041–1047. doi: 10.1002/j.1460-2075.1986.tb04320.x

Schubot, F. D., Chen, C. J., Rose, J. P., Dailey, T. A., Dailey, H. A., and Wang, B. C. (2001). Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein Sci.* 10, 1980–1988. doi: 10.1110/ps.11201

Shahid, A. A., Rao, Q. A., Bakhsh, A., and Husnain, T. (2012). Entomopathogenic fungi as biological controllers: new insights into their virulence and pathogenicity. *Arch. Biol. Sci.* 64, 21–42. doi: 10.2298/ABS1201021S

Shutt, T. E., and Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* 22, 90–95. doi: 10.1016/j.tig.2005.11.007

Sievers, F., and Higgins, D. G. (2021). "The clustal omega multiple alignment package," in *Multiple Sequence Alignment. Methods in Molecular Biology*, Vol. 2231, Ed. K. Katoh (New York, NY: Humana), 3–16. doi: 10.1007/978-1-0716-1036-7_1

Slijepcevic, P. (2021). Serial endosymbiosis theory: from biology to astronomy and back to the origin of life. *Biosystems* 202:104353. doi: 10.1016/j.biosystems.2021.104353

Sologub, M., Litonin, D., Anikin, M., Mustaev, A., and Temiakov, D. (2009). TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell* 139, 934–944. doi: 10.1016/j.cell.2009.10.031

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Strasser, H., Vey, A., and Butt, T. M. (2000). Are there any risks in using entomopathogenic fungi for pest control, with particular reference to the bioactive metabolites of *Metarhizium*, *Tolypocladium* and *Beauveria* species? *Biocontrol Sci. Technol.* 10, 717–735. doi: 10.1080/09583150020011690

Swofford, D. L. (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version, 4.0, Beta 10*. Sunderland, MA: Sinauer Associates. doi: 10.1111/j.0014-3820.2002.tb00191.x

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkaa1074

Theelen, B., Christinaki, A. C., Dawson, T. L. Jr., Boekhout, T., and Kouvelis, V. N. (2021). Comparative analysis of *Malassezia furfur* mitogenomes and the development of a mitochondria-based typing approach. *FEMS Yeast Res.* 21, foab051. doi: 10.1093/femsyr/foab051

Tracy, R. L., and Stern, D. B. (1995). Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr. Genet.* 28, 205–216. doi: 10.1007/BF00309779

Typas, M. A., and Kouvelis, V. N. (2012). "Assessing genotoxic effect s of microbial products," in *Beneficial Microorganisms in Agriculture, Food and the Environment: Safety Assessment and Regulation*, eds I. Sundh, A. Wilcks, and M. S. Goettel (Wallingford: CAB International), 256–274. doi: 10.1079/9781845938109.0256

Velazquez, G., Sousa, R., and Brieba, L. G. (2015). The thumb subdomain of yeast mitochondrial RNA polymerase is involved in processivity, transcript fidelity and mitochondrial transcription factor binding. *RNA Biol.* 12, 514–524. doi: 10.1080/15476286.2015.1014283

Vey, A., Hoagland, R. E., and Butt, T. M. (2001). "Toxic metabolites of fungal biocontrol agents," in *Fungi as Biocontrol Agents*, eds T. M. Butt, C. W. Jackson, and N. Magan (Oxon: CABI Publishing), 311.

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* 54, 5–6. doi: 10.1002/cpbi.3

Werner, F. (2007). Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* 65, 1395–1404. doi: 10.1111/j.1365-2958.2007.05876.x

Werner, F. (2012). A nexus for gene expression—molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* 417, 13–27. doi: 10.1016/j.jmb.2012.01.031

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9, 85–98. doi: 10.1038/nrmicro2507

Xiao, G., Ying, S. H., Zheng, P., Wang, Z. L., Zhang, S., Xie, X. Q., et al. (2012). Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci. Rep.* 2:483. doi: 10.1038/srep00483

Yang, X., Chang, H. R., and Yin, Y. W. (2015). Yeast mitochondrial transcription factor Mtf1 determines the precision of promoter-directed initiation of RNA polymerase Rpo41. *PLoS One* 10:e0136879. doi: 10.1371/journal.pone.0136879

Yildiz, G., and Ozkilinc, H. (2021). Pan-mitogenomics approach discovers diversity and dynamism in the prominent brown rot fungal pathogens. *Front. Microbiol.* 12:647989. doi: 10.3389/fmicb.2021.647989

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

Conflict of Interest: The 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.

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership