scientific reports

Check for updates

OPEN

Co-evolution of large inverted repeats and G-quadruplex DNA in fungal mitochondria may facilitate mitogenome stability: the case of *Malassezia*

Anastasia C. Christinaki[1,2,6], Bart Theelen[2,6], Alkmini Zania[1,2], Selene Dall' Acqua Coutinho[3], Javier F. Cabañes[4], Teun Boekhout[2,5] & Vassili N. Kouvelis[1]✉

Mitogenomes are essential due to their contribution to cell respiration. Recently they have also been implicated in fungal pathogenicity mechanisms. Members of the basidiomycetous yeast genus *Malassezia* are an important fungal component of the human skin microbiome, linked to various skin diseases, bloodstream infections, and they are increasingly implicated in gut diseases and certain cancers. In this study, the comparative analysis of *Malassezia* mitogenomes contributed to phylogenetic tree construction for all species. The mitogenomes presented significant size and gene order diversity which correlates to their phylogeny. Most importantly, they showed the inclusion of large inverted repeats (LIRs) and G-quadruplex (G4) DNA elements, rendering *Malassezia* mitogenomes a valuable test case for elucidating the evolutionary mechanisms responsible for this genome diversity. Both LIRs and G4s coexist and convergently evolved to provide genome stability through recombination. This mechanism is common in chloroplasts but, hitherto, rarely found in mitogenomes.

Mitochondria are semi-autonomous organelles with their own genome, providing multiple vital processes, such as energy generation in the form of adenosine triphosphate (ATP) to the eukaryotic cell[1]. Fungi are lower eukaryotes found in every habitat and involved in all aspects of life and they act as decomposers, parasites, pathogens, or symbionts to other eukaryotic organism[2]. Fungal mitogenomes generally contain 14 protein-coding genes (*atp*6, *atp*8, *atp*9, *cob*, *cox*1-3, *nad*1-6 and *nad*4L), two ribosomal RNA (rRNA) genes (*rns* and *rnl*) and a variable number of transfer RNA genes (*trn*s), similarly to the mitogenomes of metazoa[1]. They may also encode other accessory genes like the ribosomal protein (*rps*3 or *var*1) and the RNA component for RNaseP (*rnp*B)[3,4]. Fungal mitochondrial genomes present a remarkable size variation ranging from 11,223 bp (*Hanseniaspora pseudoguilliermondii*)[5] to 332,165 bp (*Golovinomyces cichoracearum*, NCBI Acc. No: NC_056148) and a variable gene order, which render them assets for studying fungal evolution[5–7]. This diversity of the mitogenomes is the result of the absence or presence of intergenic spacers, duplications, repeats, and insertions of plasmid components or other elements[4]. Conserved large inverted repeats (LIR), although scarcely found, have been linked to mitogenome regulation[8]. Similarly, G-quadruplexes (G4) in mitogenomes are involved in the regulation of important biological processes, such as replication, altered gene expression, recombination, double-stranded break (DSB) formation, telomere regulation and genome instability[9,10], as shown in the yeast *Saccharomyces cerevisiae*[8,11].

Among the fungal species examined for their mt genome diversity, the basidiomycetous yeast *Malassezia* is an intriguing study case due to the mode of life of these yeasts[12]. It is the most abundant fungal genus on healthy human skin[13] but may also cause various skin diseases such as atopic dermatitis, seborrheic dermatitis, dandruff,

[1]Section of Genetics and Biotechnology, Department of Biology, National and Kapodistrian University of Athens, Panepistimiopolis, 15771 Athens, Greece. [2]Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [3]Laboratory of Molecular and Cellular Biology, Veterinary Medicine, Paulista University, São Paulo, Brazil. [4]Veterinary Mycology Group, Department of Animal Health and Anatomy, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain. [5]College of Science, King Saud University, Riyadh, Saudi Arabia. [6]These authors contributed equally: Anastasia C. Christinaki and Bart Theelen. ✉email: kouvelis@biol.uoa.gr

and pityriasis versicolor[14,15]. In recent years, *Malassezia* has increasingly been implicated in health and disease beyond the skin: as an underestimated cause of *Malassezia* bloodstream infections (BSIs) in immunocompromised patients and neonates[16], as resident of the healthy human gut mycobiota[17], but it has also been associated with Crohn's disease[18], colorectal cancer[19,20], promoting pancreatic oncogenesis[21], and is implicated in various respiratory diseases[22]. Several *Malassezia* species may also cause cutaneous pathologies in a variety of animals[12]. Direct sequencing approaches investigating various environmental sources, such as insects, nematodes, sponges, corals, soils, deep-sea vents and subglacial habitats, showed that this yeast may be ecologically more diverse than believed until now[12,23].

Malassezia is lipid dependent and many species have slow and fastidious growth[24]. Due to culturing limitations, there are challenges to obtain living cells from complex clinical and environmental sources to accurately represent the role of *Malassezia* species and to better understand their evolution and the underlying mechanisms in various diseases and ecologies. With recent advancements in genomics, nuclear genomes are now available for all currently described species[24–27] providing valuable resources for examining functional aspects and the evolutionary trajectory of this genus. However, only a limited number of studies have covered mitochondrial aspects and their contribution the evolution of the genus. Recently, *Malassezia sympodialis* and *Malassezia furfur* mitogenomes revealed LIRs[3,26,28].

In this study, the mitogenomes of 11 *Malassezia sp.* were de novo sequenced and annotated in order to expand the analysis to all species of this genus. In detail, the first comprehensive analysis of the mitochondrial genomes of all 18 currently described *Malassezia* species plus two putative new ones provided insights into *Malassezia* mt genome organization and evolution. Furthermore, the potential roles of LIRs, and G4s are discussed since these molecular elements seem to coexist and possibly could act as a unit in genome recombination and stabilization. This work raised the question whether these two features are functionally correlated and evolved convergently.

Results

General mitogenomic features reveal significant size variation and diversity. To date the mitochondrial genomes of nine *Malassezia* species are available in GenBank but only the ones of *M. furfur* and *M. sympodialis* have been described in detail[26,28]. In this study, newly acquired genomes of the remaining 11 species, including a hitherto undescribed species from bat, where sequenced, and their mitogenomes were assembled and annotated. In total, 28 mitogenomes were analyzed, representing all known *Malassezia* species including multiple strains, when available, for an in-depth pan-*Malassezia* comparative mitogenomic approach (Table 1).

Mitogenomes varied in size between 28,009 and 63,712 bp, with the one of *Malassezia japonica* being the smallest and the genome of *Malassezia cuniculi* the largest. This size variation mostly resulted from differences in number and size of introns, the presence of diverse LIRs and the size of intergenic regions (Suppl. Table S1). The number of introns varied between zero for *M. japonica* and two strains of *M. globosa*, and ten for hybrid *M. furfur* strain CBS 7019 with a total intron size ranging from zero to 9172 bp (for *M. obtusa*). All introns were of group I type, located in the *rnl*, *cox1* and *cob* genes. Several introns hosted putative LAGLIDADG homing endonuclease genes (HEGs), with the exception of the first intron (ID subtype) in the *cob* gene of *M. pachydermatis* strain CBS 1879, which contained a GIY-YIG homing endonuclease gene (Suppl. Table S2).

The total number of nucleotides allocated to intergenic regions varied between 13,612 bp for *M. obtusa* and 47,344 bp for *M. cuniculi*. The GC-content was low as expected for mitochondrial genomes, approximately 30%. The mitogenomes for all analyzed strains contained the expected standard 14 protein-coding genes, two rRNA genes, between 22 and 32 tRNAs, and the protein coding *rps*3 gene, with *M. slooffiae* and *M. cuniculi* containing two copies of this gene. *M. slooffiae* uniquely possessed the gene for the RNA subunit of mitochondrial RNase (*rnp*B). Most *Malassezia* mitogenomes contained a LIR with a variable number of duplicated genes (Suppl. Table S1).

For some species, mitogenomes for multiple strains were available, reanalyzed, and assessed, showing a variable level of intraspecies mitogenomic differences mostly resulting from the diversity of their intergenic regions and LIRs, and their intron abundance. For example, of the three *M. globosa* strains, the mitogenomes of the CBS 7966 and CBS 7874 do not possess any introns, but CBS 7991 contains three introns.

Up to now an intraspecies mitogenome variability was previously analyzed for 20 M. *furfur* strains[3]. The authors reported a mitogenome size variability ranging from 45,715 to 49,317 bp, mostly resulting from variation in intron abundance and variable intergenic regions[3]. In the present study the mitogenome of CBS 9595 was selected and analyzed, as this turned out to be an important parental reference strain in a study focusing on *M. furfur* hybridization events[28].

Mitochondrial phylogeny. Phylogenetic relationships were determined considering the 15 mt protein coding genes, with *Ustilago maydis* used as outgroup (Fig. 1). Three main phylogenetic clades could be determined: clade I, containing *M. arunalokei*, *M. restricta*, and *M, globosa*; clade II, with *M. caprae*, *M. dermatis*, *M. sympodialis*, *M. equina*, *M. nana*, and *M. pachydermatis*; clade III, containing *M. brasiliensis*, *M. furfur*, *M. obtusa*, *M. psittaci*, *M. yamatoensis*, *M. japonica*, and *M. vespertilionis*; and finally, the basally positioned species *M. slooffiae* and *M. cuniculi*, together with strain CBS17886 which represents a putative new species. In current mitogenomic phylogeny, *M. furfur* CBS 9595 was positioned slightly deviating from the other four *M. furfur* strains with separate positioning, and in a cluster also containing *M. brasiliensis*. Based on the concatenated 15 protein coding genes (PCGs), sequence similarity between CBS 9595 and the other four *M. furfur* strains was 98.3–98.4% compared to 99.7–99.9% among the four closer related strains, corroborating previous indications that *M. furfur* may comprise two closely related species[28] (Suppl. Table S3).

#	Species	Strain	Origin	Source	NCBI Acc. No	mtDNA (bp)	mt phylogenetic clades
1	*M. arunalokei*	CBS 11387	India	Scalp, SD patient	**ON585844**	38143	I
2	*M. restricta*	KCTC 27527	Korea	Dandruff	NC_039455.1	38720	I
3	*M. restricta*	CBS 7877	UK	Human	KY911093.1	38499	I
4	*M. globosa*	CBS 7966	UK	Skin, PV patient	KY911088.1	34689	I
5	*M. globosa*	CBS 7874	UK	Scalp, dandruff	KY911087.1	34808	I
6	*M. globosa*	CBS 7990	UK	Human	KY911089.1	38672	I
7	*M. caprae*	CBS 10434	Spain	Ear, goat	**ON585846**	41692	II
8	*M. dermatis*	CBS 9169	Japan	Skin lesions, atopic patient	**ON585848**	38991	II
9	*M. sympodialis*	ATCC 44340	N/A	N/A	KY911095.1	38614	II
10	*M. sympodialis*	ATCC 42132	N/A	Skin of TV patient	HF558646.1	38622	II
11	*M. sympodialis*	ATCC 96806	UK	Healthy skin	KY911096.1	38538	II
12	*M. equina*	CBS 12830	Spain	Anus, horse	**ON585849**	45405	II
13	*M. nana*	CBS 9557	Japan	Otitis externa, feline	**ON585850**	42091	II
14	*M. pachydermatis*	CBS 1879	Sweden	Otitis externa, ear, dog	KY911092.1	35527	II
15	*M. brasiliensis*	CBS 14135	Brazil	Beak, parrot	**ON585845**	60300	III
16	*M. furfur*	CBS 9595	Greece	SD, back, human	**ON585854**	48295	III
17	*M. furfur*	CBS 1878	N/A	Dandruff, human	KY911081.1	48161	III
18	*M. furfur*	CBS 7019	Finland	PV, trunk, 15-year-old girl	KY911083.1	49305	III
19	*M. furfur*	CBS 14141	France	Catheter, blood, human	KY911086.1	48933	III
20	*M. furfur*	CBS 7982	France	Skin of ear, healthy human	KY911085.1	47903	III
21	*M. obtusa*	CBS 7876	UK	Human	KY911091.1	60147	III
22	*M. psittaci*	CBS 14136	Brazil	Beak, parrot	**ON585852**	46165	III
23	*M. yamatoensis*	CBS 9725	Japan	SD	KY911098.1	41233	III
24	*M. japonica*	CBS 9431	Japan	Skin, healthy Japanese woman	KY911090.1	28009	III
25	*M. vespertilionis*	CBS 15041	USA	Wing, skin, bat (Myotis septentrionalis)	**ON585851**	59251	III
26	*Malassezia sp.*	CBS 17886	USA	Bat (Lasionycteris noctivagans)	**ON585853**	41053	Basal
27	*M. slooffiae*	CBS 7956	France	Ear, healthy pig	KY911097.1	40575	Basal
28	*M. cuniculi*	CBS 11721	Spain	External ear canal, healthy rabbit	**ON585847**	63712	Basal
29	*Ustilago maydis*	521	USA	Zea mays	NC_008368.1	56814	Outgroup

Table 1. *Malassezia* species and strains used for the mitogenomic comparative analysis. Their origin, source is provided along with the NCBI Accession number and the size (in bp) of their mitogenome. The mitochondrial phylogenetic clade in which each species belong is also indicated. Bold lettering indicates the newly acquired mt genomes, sequenced and annotated in this work. *NA* non-available, *PV* pityriasis versicolor, *SD* seborrheic dermatitis and *TV* tinea versicolor.

Mitochondrial synteny analyses suggest evolution towards fixated genome organization. Mitogenomic gene shuffling between species was observed, and only two groups of species presented a similar syntenic configuration (Fig. 1). Species of phylogenetic clades I and II, except for *M. nana*, showed identical gene order indicating late evolutionary acquired genome conservation. Some LIR variation existed for *M. nana* and *M. restricta* strain KCTC 27527 (see next section). The other group with identical synteny consisted of the closely related species *M. brasiliensis*, *M. furfur*, and *M. obtusa* from phylogenetic group III. One deviation in genome organization could, however, be observed in *M. obtusa*, for which the *trn*G-*nad*2-*nad*3 gene cluster was positioned in an inverted orientation compared to *M. brasiliensis* and *M. furfur*. The remaining seven species retained unique genome organizations with only partially shared synteny. Among all *Malassezia* species, six syntenic blocks could be distinguished: *trn*M-*trn*E-*trn*T, *nad*2-*nad*3, *cox*3-*nad*6, *trn*V-*cox*2-*nad*1, *trn*D-*nad*4L-*nad*5, and *trn*A-*trn*F-*trn*Y-*trn*S (Fig. 2, i–vi). If *M. cuniculi* is excluded, three more syntenic blocks could be identified (i.e., *rnl*-*trn*P-*trn*N, *trn*I-*cob* and *trn*S-*cox*1-*atp*8) (Fig. 2, vii–ix). Syntenic diversity was illustrated by the transposition and inversion of gene blocks and single genes among species. These events most probably occurred among the earlier divergent species, with the *M. cuniculi* mt genome containing the most differing genome organization. As phylogenetically basal species possessed a less organized gene order among them compared to the more evolved clusters, an evolutionary drive towards fixated genome organization seems to exist.

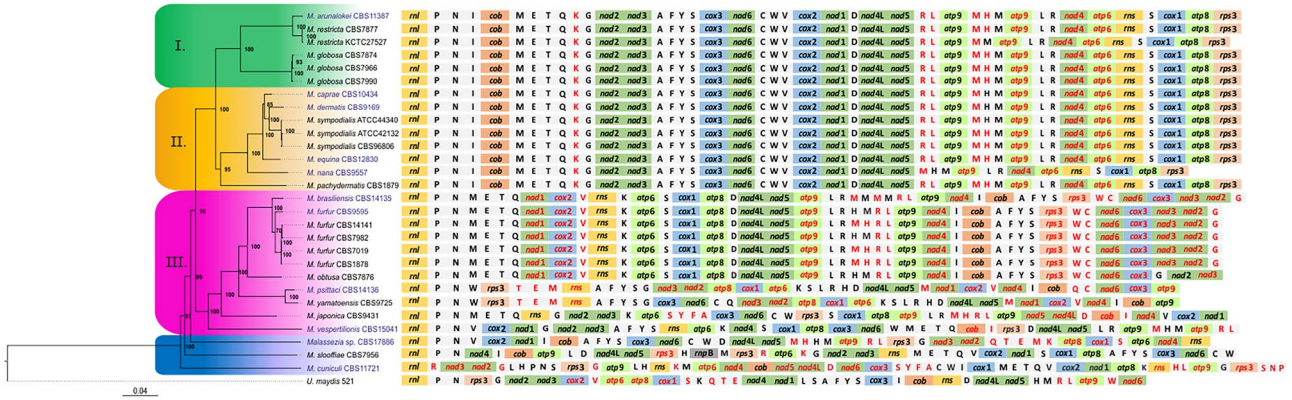

Figure 1. Phylogeny and synteny of the mitogenome of 28 *Malassezia* strains representing all currently known species. The phylogenetic tree was produced using Bayesian Inference based on the concatenated amino acid matrix of the 15 mitochondrial PCGs. Numbers at the nodes of the tree indicate posterior probability (PP) values. The species *Ustilago maydis* strain 521 has been used as an outgroup. Shading of phylogenetic clusters is shown as green, yellow, purple, and blue for clades I, II, III and the basal species, respectively. Blue text indicates the taxa whose mt genome is presented for the first time in this study. Additionally, synteny is provided in all cases starting with *rnl* gene. The genes which are located on the complementary reverse strand are marked with red color. Amino acid symbols represent the corresponding tRNA gene (e.g., P for *trn*P). Genes referring to subunits of NADH complex, apocytochrome b, cytochrome c oxidase complex and ATP synthase complex are shown in green, orange, blue, light green respectively.

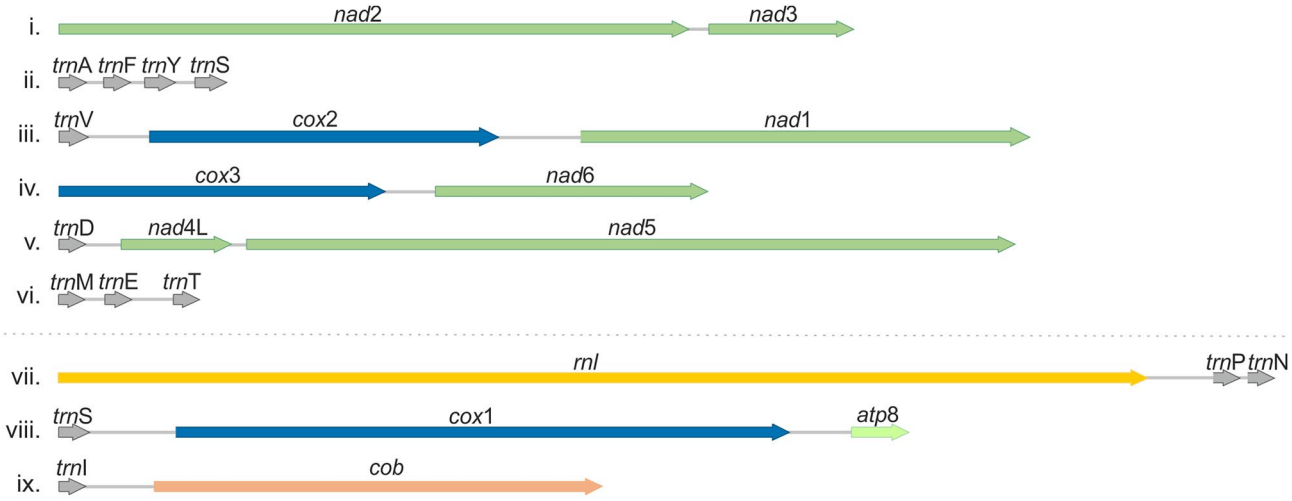

Figure 2. The main syntenic units of *Malassezia* mitogenomes. The first six units can be found in all *Malassezia* spp. (i.e., i–vi) while the last three in all except *M. cuniculi* (i.e., vii–ix). Arrows indicate gene transcription direction.

Long inverted repeats (LIRs) contribute to the conserved synteny of the more recently evolved phylogenetic clades. Nearly all *Malassezia* species contained a LIR varying in size between 2098 bp for *M. japonica* and 10,698 bp for *M. brasiliensis* (Suppl. Fig. S1). Variation in LIR assembly resulted from the presence or absence of *trn* genes in the repeat unit and in the intra-IR region, the orientation or combinations of genes, and finally from variation in the intergenic regions (Fig. 3).

The early diverged mt genomes of *M. cuniculi* and *M. slooffiae* evolved independently, resulting in unique synteny and LIR composition. In *M. cuniculi* a duplicated region composed by *trn*P-*trn*N-*trn*S-*rps3*-*trn*G-*atp*9-*trn*L-*trn*H-*rns*-*trn*K was present with the second copy positioned in an inverted orientation and separated by a long genomic stretch containing *rnl*-*trn*R-*nad*3-*nad*2-*trn*G-*trn*L-*trn*H (Fig. 3). Thus, it should probably be disqualified as a LIR but attributed to a simple inverted duplication. This duplication may have been the result of a recombination event and it facilitated the existence of duplicated *rps*3 and *rns* gene copies. In *M. slooffiae* the repeat unit contained the gene *rps*3 with genes *trn*H, *rnp*B, *trn*M located in the intra-IR region. The protein coding gene *rnp*B is only present in the mitogenome of *M. slooffiae* and it seems that this gene was introduced with the insertion of the LIR.

In CBS17886, representing a yet undescribed *Malassezia* species, a 3.7 kb fragment including the genes *trn*M and *trn*H was observed comprising the first evolutionary step for the formation of laterally evolved LIRs in all

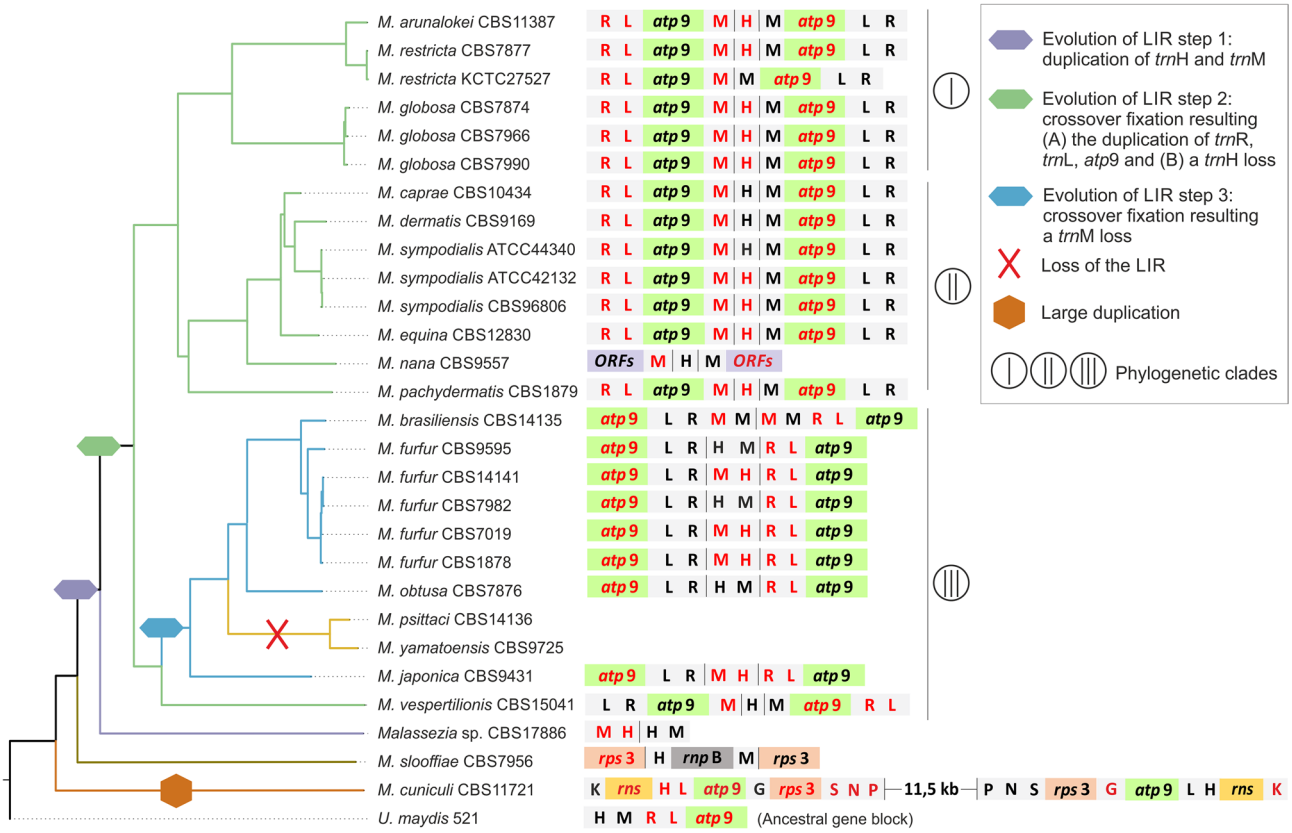

Figure 3. Schematic representation of *Malassezia* LIR evolution in relation to phylogeny. LIRs synteny is presented next to the respective phylogenetic clade. The intra-IR region can be seen in between black vertical lines. The genes located on the complementary reverse strand are indicated in red coloring and amino acid symbols represent the corresponding tRNA gene (e.g., H for *trn*H). The main evolutionary steps are shown as different colored symbols of the nodes and clades of the tree. Latin numbers in circles indicate the three different phylogenetic clades. The ancestral gene block of *Ustilago maydis* is provided.

other *Malassezia* species. Genes included in *Malassezia*'s LIRs were also found in the *U. maydis* mitogenome as an ancestral gene block composed by *trn*H-*trn*M-*trn*R(inverted)-*trn*L(inverted)-*atp*9(inverted). Phylogenetic clades I and II differed from clade III in their LIR and intra-IR content. In clades I and II, *trnM-atp*9(inverted)-*trnL-trnR* constituted the LIR and the intra-IR included the *trnH* gene. In clade III (with the exception of *M. vespertilionis*), the LIR included only *trnR*(inverted)- *trnL*(inverted)-*atp*9, and the intra-IR consisted of *trnH-trnM*. Two species of clade III, *M. yamatoensis* and *M. psittaci*, lacked the LIR entirely, which may have been lost in a more recent evolutionary event, as both species clustered together in a separate clade (Fig. 3).

In *M. furfur* and *M. sympodialis,* intraspecies LIR variation was also detected and mainly resulting from different orientations of the intra-IR fragment. Interestingly, one of the *M. restricta* strains lacked the intra-IR region.

G-quadruplex (G4) motifs in *Malassezia* mitogenomes are predominantly distributed in the LIRs. Potential G4 formation was identified in all examined mitogenomes when the threshold of 1.2 and the window size of 25 (default parameters) were used. The number of found G4s varied from one in *M. yamatoensis* to 37 in *M. obtusa*, without significant correlation with the mitogenome size and the GC content (Suppl. Table S4, Suppl. File S1). G4s are mainly distributed within the LIRs, with few exceptions. *M. globosa* was the only species without any G4s in the LIR (Fig. 4). In the *M. yamatoensis* and *M. psittaci* mitogenomes, which do not possess LIRs, G4s were located within the intergenic regions. Interestingly, only phylogenetic clades I and II incorporate G4DNA motifs in *trn*P, and their distribution is conserved within each of these clades. Potential G4s were also observed (a) in *cox*1 third intron (group-IB) of *M. sympodialis* and *U. maydis*, (b) in *atp*9 of the early diverging species *M. slooffiae* which is not part of its unique LIR, and (c) in the *trn*K of *M. caprae.* G4 analysis showed that some predicted G4s have only three runs of Gs instead of the four needed, with at least two consecutive Gs. This could be explained by (a) the potential presence of bulges found in adjacent single Gs and (b) the presence of G-runs upstream and downstream of the predicted G4s. For this reason, the analysis was repeated with the same threshold (1.2), but with larger window size (30). Interestingly, the predicted G4s for this latter analysis were only located in LIRs (94.5%) and intergenic regions (5.5%) (Fig. 4, Suppl. Table S4). With these stricter parameters, the species *M. globosa*, *M equina* and *M. yamatoensis* did not possess potential G4s. When comparing the results of G4 analyses with different window sizes, it became evident that G4s with scores more than 1.2 could result in both 25 and 30 window size, depending on the species and its mitogenome sequence

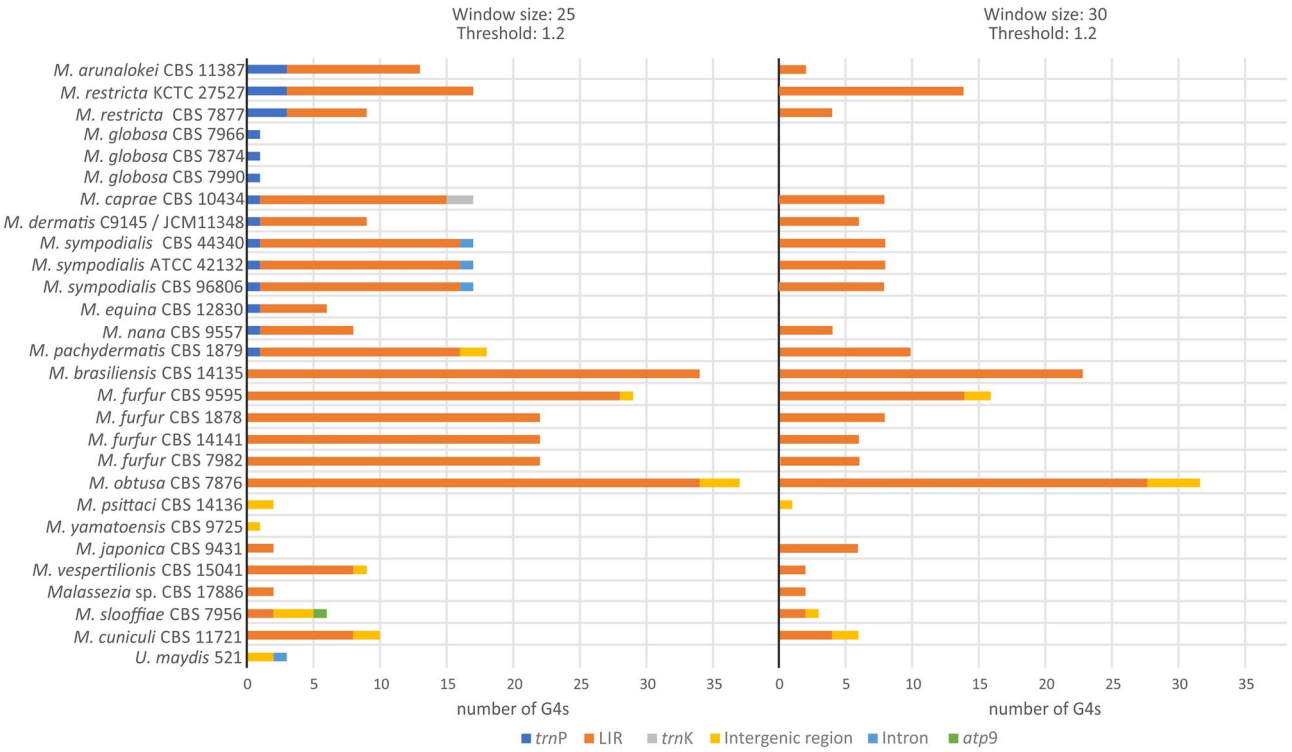

Figure 4. The G-quadruplex distribution in *Malassezia* mitogenomes. Bar graphs of the predicted G4s with threshold 1.2 and window size 25 (left) or window size 30 (right). The G4s located in the *trn*P, LIRs, *trn*K, intergenic regions, introns and *atp*9 gene are shown with blue, orange, grey, yellow, light blue and green color, respectively.

(Table 2). Among all predicted G4s with both parameters, the loop length varied from one to 18 bp, and G4 sequence conservation was only detected at the intra-species level (Suppl. Table S4).

Discussion

The mitochondrion is a vital organelle for cell growth and survival of eukaryotic cells including fungi, but recent research has also implicated mitochondrial functions in the ability for pathogenic fungi to cause disease[29,30]. Mitochondrial genomes may also provide useful insights into fungal evolution as previously shown[4,5]. In general, whole genome sequencing projects focus on the nuclear genomes and thus far, only 192 out of 1,189 available fungal mt genomes belong to Basidiomycetes (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/). However, species of Basidiomycetes present a complex mitogenomic gene organization with both ancestral and newly acquired properties, rendering them excellent model organisms for studying fungal evolution. Additionally, the medically relevant genus *Malassezia* embodies scarcely observed genetic elements, like LIRs, in their mitogenomes[3,28]. In this study, the mitogenomes of 11 *Malassezia* species along with publicly available mitogenome data (totaling 28 mt genomes, representing 18 described species and two putative new species) were utilized in a comparative genome analysis of all species of the genus *Malassezia*, in order to unravel the evolution of its mitogenome.

As expected, all *Malassezia* mt genomes contain 14 protein-coding genes that are present in nearly all fungi and metazoa, plus the *rps*3 gene which is present only in ca. 65% of Basidiomycetes[31]. *Malassezia* mitogenomes demonstrate substantial variation among species, mostly resulting from variation from intron size and frequency, intergenic regions, and LIR diversification, leading to mt genome sizes ranging from 28 to 64 kb.

A recent study, presenting a whole nuclear genome based phylogenetic tree, identified a similar main clustering as the mitochondrial phylogeny of this work, with some deviations and in some cases slightly different species positions withing the main clusters. Most noteworthy, in the nuclear phylogeny, *M. cuniculi* and *M. slooffiae* are also basally positioned but seem to have a sister relationship, whereas in the mitochondrial phylogeny, *M. cuniculi* is ancestral to *M. slooffiae*. Such discrepancies between mitochondrial and nuclear phylogenies have been described before and may result from different selection pressures on both genomes[32] and also showcase the added value of concatenated mitochondrial PCG matrices in fungal phylogenetics, as was recently also shown for the subphylum Saccharomycotina of Ascomycota[5].

In fungi, variation in gene order is immense[33–35]. Even when only considering the genus *Malassezia*, a diversity in mt genome organization was observed (Fig. 1). Only six shared syntenic blocks exist (Fig. 2), of which *nad*2-*nad*3 seems to form a syntenic block in most fungal species[36]. The only other two genes that generally exist next to each other in fungi are *nad*4L-*nad*5 and it has been suggested that the variations in gene order is most likely

Species/strain	G4 Length	Sequence (5′-3′)	G4Hunter score (abs)	Location	Window size
M. arunalokei CBS11387	40	TATAGGGGTTACTCTAATGGGGTTTTAGGGTAATGGGTTT	1,2	LIR	30
M. restricta KCTC 27527	25	GGGTAGGTAGGTGGGTATAGTAGAG	1,2	LIR	25
M. restricta CBS7877	32	ACAGAGGGGAAGTATAGACAGAGGGGAAGTGG	1,219	LIR	25
M. caprae CBS10434	30	GTGGGTGGGGTTAATGGATGGGATGTAGTG	1,4	LIR	30
M. dermatis C9145 / JCM11348	31	GGGGGGAAGGTAAGATAAATGGCTAAATGGG	1,323	LIR	30
M. sympodialis CBS44340	33	TGGGTTAGGTAGGGAATTAGTAGAGGGTAGGGT	1,273	LIR	30
M. sympodialis ATCC 42132	33	TGGGTTAGGTAGGGAATTAGTAGAGGGTAGGGT	1,273	LIR	30
M. sympodialis CBS96806	33	TGGGTTAGGTAGGGAATTAGTAGAGGGTAGGGT	1,273	LIR	30
M. nana CBS9557	30	GTAGGGGTGTATCAAGGGGCGATTTATAGG	1,233	LIR	30
M. pachydermatis CBS1879	31	GGGGAGTAAGTAAGGGTATATGGTTAAAGGG	1,29	LIR	30
M. brasiliensis CBS14135	33	AGGGATAGGATAGGATAGGGATAGGGATAGGGA	1,333	LIR	30
M. furfur CBS9595	30	GGATAGGGGATGGTATGGATACAGTATGGG	1,233	LIR	30
M. furfur CBS1878	30	GGTAAAGATTTGTAGGGTGGTAGTAAGGGG	1,2	LIR	30
M. furfur CBS14141	25	GGGGAGGAGTAGAGTATTATATGGG	1,28	LIR	25
M. furfur CBS7982	25	GGGGAGGAGTAGAGTATTATATGGG	1,28	LIR	25
M. obtusa CBS7876	25	GGGGTATAGTATGTGGTATAGTGGG	1,28	LIR	25
M. psittaci CBS14136	29	AAGGGTTGAAAGGGTGGTTGAAAGGGTGG	1,276	*trn*T-*trn*E intergenic region	25
M. yamatoensis CBS9725	25	GGGATAGTTGGGATAGTTGGGATAG	1,2	*cob-atp*9 intergenic region	25
M. japonica CBS9431	35	AATGGGGGTTGAAAATAAGTAGTAAGAGGGGTAGG	1,257	LIR	30
M. vespertilionis CBS15041	27	GGGTTTGGGTTTGTTGTAGGGTTTGGG	1,407	LIR	25
Malassezia sp. CBS17886	25	TATAGGGGTTTTTTACTCGAAGGGG	1,24	LIR	30
M. slooffiae CBS7956	25	GGAGCTGGGGTAGGAGTAGGAATTG	1,2	*atp*9	25
M. cuniculi CBS11721	25	GGGATGGGGTAAAATTTATAGGTAG	1,2	repeat region	25
U. maydis 520	25	GAGGAAATGTGAAGGGGGAGTAGTG	1,2	*atp*8-*cox*1 intergenic region	25

Table 2. Predicted G4s with the highest score in G4hunter tool for each *Malassezia* species. G4 length, sequence, score and location on the mitogenome is provided. G4 residues are colored based on their contribution to the final score, i.e., dATPs and dTTPs are neutral (black color), dGTPs are positive (red color) and dCTPs negative (blue color). The window size, in which the best G4 was predicted, is also indicated. *M. globosa* and *M. equina* are not included since the score of the predicted G4s did not exceed 1.2.

the result of recombination[5,34]. Gene organization seems to be more discrepant between basal species and has a tendency to stabilize in the lately evolved phylogenetic clades.

All but two *Malassezia* species contain a LIR of variable size and content (Fig. 3). Even the mitogenome of the earlier diverging species *M. cuniculi*, contains a unique inverted duplication (Fig. 3). Another, evolutionarily independent species-specific duplication happened in *M. slooffiae,* which is the only species that additionally contains the *rnp*B gene which is uncommon among Basidiomycetes. This gene encodes the RNA subunit of mitochondrial RNaseP, an enzyme necessary for the maturation of the tRNAs[34]. It has been shown that *rnp*B is a functional gene in Ascomycetes and the protists *Reclinomonas americana* and *Nephroselmis olivacea*, and it is considered an ancestral element which was probably lost in most fungal species during evolution[37]. However, the existence of this gene in the intra-IR region of *M. slooffiae* may suggest a evolutionarily more recent addition to the mitogenome due to the LIR mobility. The LIRs of the remaining species seem to be the result of a continuous evolutionary process of three steps, starting with the duplication and inversion of the ancestral syntenic unit *trn*H-*trn*M in the mitogenome of *Malassezia* sp. CBS17886 (Fig. 3, step 1). As a second step of the process, the duplication of ancestral syntenic unit *trn*R-trnL-*atp*9, also found in *U. maydis* (NCBI Acc No. NC_008368.1) and *Pseudozyma* sp. (NCBI Acc No. MK714018.1), and the loss of the one *trn*H copy led to the formation of the LIRs in phylogenetic clades I, II and in *M. vespertilionis.* Subsequently, in the other species of clade III an inversion of the previous syntenic unit is observed along with the deletion of one of the two *trn*Ms, resulting in an intra-IR region containing only *trn*M-*trn*H. In agreement with previous studies, the mechanism might have been recombination[38] which in this case resulted in a cross-over fixation, leading tothe loss of *trn*M. Interestingly, between closely related species with comparable LIR composition in clades I, II and III, the intra-IR is detected in normal or inverted orientation. Studies focusing on intra-species variation confirmed the presence of multiple

mt genome rearrangements in various strains of one species, resulting from recombination between LIRs[3,26]. This could possibly apply for all *Malassezia* species but requires further exploration. The LIR variability in *Malassezia* is most probably the main course for the diverse synteny seen within the genus. This hypothesis is in agreement with the work of Liu and colleagues (2020) who demonstrated the presence of two different mtDNA genotypes in the same cell of *Agrocybe aegerita*, as a result of intramolecular recombination[39]. Up to date, LIRs have only been described for a few fungal genera, including several *Candida* species[28,38,40], *Saccharomyces cerevisiae*[41], *Hanseniaspora* sp. (i.e., *Kloeckera africana*)[42] and several species of Agaricales[43]. Mito-LIRs can be found in a certain array of "lower" eukaryotes, like the oomycetes *Achlya ambisexualis*[44] and the protozoan *Tetrahymena pyriformis* in which the mitogenome is linear[45].

Previous studies showed that LIRs are recombination hotspots leading to the initiation of replication in *Candida albicans*[38]. A comparative study exploring the mt genomes of eight *Candida* species determined that LIRs may be involved in the so-called flip-flop recombination resulting in genome isomers and it was demonstrated that LIRs are involved in conversions between circular and linear mitochondrial topologies and in extent they may play a role in replication[40]. A possible mode of LIR introduction into mitochondria is plasmid integration[46], although this study does not confirm this mechanism of origin for *Malassezia*, since no DNA polymerase gene or other plasmid related elements were found in all species examined. The lack of LIRs in two *Malassezia* species indicates that replication is not performed due to the LIRs. Finally, the LIR variation with its expansion and/or contraction within the genus *Malassezia* suggests that it is an ongoing process with back-and-forth results as observed in the different species.

In general, fungal mitogenome analyses present considerable genome size variation, which until now has been contributed to their sequence alterations resulting from mutation and recombination[5,47]. On the other hand, in plants and more specifically in chloroplast genomes, size diversity may be explained by sequence complexity as well as by the amount of repeated DNA[48]. Chloroplast genomes (cpDNA) usually present a commonly found structure of two single regions divided by LIRs[49] allowing cpDNA to evolve under strong constraint—either mechanistic or selective—in order to maintain a compact, largely genic genome[48]. In fungi, the lack of LIRs in the majority of the species indicates a free evolution of the mitogenomes with respect to their size, with other proposed underlying mechanisms[34,50].

Additionally, the existence of inverted repeats is strongly related to the high number of G4s in the *S. cerevisiae* mitogenome[8,11]. In the current study, a similar observation is made for the mitogenomes of the genus *Malassezia* (Fig. 4). In species which do not possess LIRs, G4 formation can occur in intergenic regions, and in some cases in *trn* genes or within introns (Fig. 4). Interestingly, setting the window size to 25, potential G4 structures were overlapping with the *trn*P gene in *Malassezia*'s later diverging phylogenetic clades I and II. This phenomenon may be additional evidence to the hypothesis of *trn* genes acting as recombinational hotspots for mitogenome gene shuffling[5,51]. Moreover, this distribution contradicts with the respective G4 analysis in *S. cerevisiae*, where G4s did not overlap with *trn*s[11]. G4 DNA structures have already been related to DSBs[9,52]. Their location within the LIRs, which have been implicated in recombination events whenever DSBs exist, offer a repair mechanism and contribute to genome stability, despite the G4 DNA implication to deleterious effects on mtDNA[52].

The following hypothesis concerning fungal mitogenome evolution is presented: In contrast to the cpDNAs, mitogenomes rarely contain cpDNA-like structures, and whenever found in fungi they are tandemly scattered, i.e., in different orders/subphyla which are also phylogenetically distant. Thus, their acquisition must be a recent event which happened several times during fungal evolution, as phylogenetically scattered species carrying LIRs in their mt genome suggest. The G4 DNA structures seem to have coevolved with the LIRs, whenever both exist. In other words, they coexisted to act as a unit performing recombination and repair to stabilize the mitogenome.

In conclusion, *Malassezia*'s mitogenome analyses showed the existence of newly acquired elements like LIRs and G4 DNAs, which converged and provided a reliable recombination-based mechanism to facilitate genome stability. This work adds to our understanding of the mechanisms that drive mitogenomic evolution.

Materials and methods

***Malassezia* species and strains.** In this study 28 *Malassezia* strains, corresponding to all known species of the genus, including a putative new species from bat, were used in the *in-silico* comparative mt genome analysis. In detail, 11 mt genomes were retrieved from in house generated whole genome sequencing data (WGS) (Table 1).

Culture and DNA extraction. *Malassezia* strains were grown on modified Dixon agar[53] for 48–96 h at 30 °C (*M. cuniculi* on Leeming and Notman agar for 7 d at 36 °C; CBS 17886 for 7 d at 24 °C), and cells were harvested into 50-mL tubes. Yeast cells were lysed, following the Qiagen genomic DNA purification procedure for yeast samples (Qiagen, Hilden, Germany), with some modifications. Lyticase incubation was performed for 2 h at 30 °C, and RNase/Proteinase incubation was performed for 2 h at 55 °C. Genomic DNA was purified using Genomic-tip 100/G prep columns, according to the manufacturer's instructions.

Whole genome sequencing. Samples were sequenced on the Oxford nanopore PromethIon platform, and libraries were prepared using the standard Native Barcoding Genomic DNA Ligation Protocol. Base calling was performed on the PromethIon (Release 21.02.7) with high accuracy setting.

Long read assembly. De novo assembly was carried out by Flye version 2.9[54]. The "–nano-raw" flag was used; genome size was set to 8 Mb and polishing iterations to 2.

Mitochondrial DNA assembly. For most strains, mitochondrial DNA was found in multiple fragments in the assembled genomes. To retrieve it in one piece, the following process was followed: (1) raw reads were mapped against the assembled genome, using minimap2 version 2.17-r941, (2) reads that mapped against chromosomal fragments were discarded; (3) remaining reads were reassembled using both Flye 2.9[54], setting "–asm-coverage" to 50, and canu 2.2[55]. Since reads were not 100% mitochondrial and the exact size of the remaining chromosomal fragments was not known, multiple assemblies took place for every strain. Genome size was set to 50 kb, 500 kb or 1000 kb, until the mtDNA was retrieved in one scaffold. Where Illumina data were available, mtDNAs were assembled using GetOrganelle v1.7.5 using recommended parameters[56]. The obtained mt scaffolds were additional validated by aligning them against known *Malassezia* mt genomes using tBLASTn/BLASTx/BLASTn[57] and were further analyzed as described below.

Mitochondrial genome annotation and comparative analyses. The mt scaffolds were annotated as follows: the protein coding and the ribosomal (rRNA) genes were identified using BLASTx and BLASTn, respectively. The tRNA genes were detected using the web-based tRNAScan-SE platform[58]. Intron characterization was performed by RNAweasel online tool[59] and the intron borders were also verified manually. The open reading frames (ORFs) were identified using ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) setting "ORF start codon to use" parameter as "ATG and alternative initiation codons", and the genetic code employed was "The Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma/Spiroplasma Code" (NCBI transl_table=4). The 11 newly annotated mt genomes of this study were deposited to GenBank (acc Nos: ON585844-ON585854). The sequence of the nine conserved mt gene blocks identified in synteny analysis, were collected, and aligned by multiple sequence alignment program MAFFT using the E-INS-i method[60,61] in order to identify the mt sequence diversity in *Malassezia* spp.

Phylogenetic analyses. Amino acid sequences as produced by the 15 mt protein coding genes (PCGs) (i.e., *atp*6, *atp*8-9, *cob*, *nad*1-6, *nad*4L, *cox*1-3 and *rps*3) were collected from the 28 *Malassezia* mt genomes for phylogenetic purposes (Suppl. File S2). *Ustilago maydis* strain 521 was used as an outgroup (GenBank Acc. No. NC_008368) in all analyses besides the phylogenetic one. The sequences for each protein were aligned by Lasergene's MegAlign v.11 program (10.1385/1-59259-192-2:71) using the ClustalW method with default settings. The concatenated dataset was created for phylogenetic purposes. The phylogenetic tree construction was performed using Neighbor Joining (NJ) and Bayesian Inference (BI) methods through PAUP4[62] and MrBayes[63] software, respectively. For NJ analyses, reliability of nodes was evaluated using 10.000 bootstrap iterations for all concatenated and individual datasets. For BI analyses, the determination of the evolutionary model, which was best suitable for each dataset, was performed using the program ProtTest (ver. 3)[64]. The BIC Information Criterion was applied, and the best nucleotide substitution model was found to be LG + I + G + F. Four independent MCMCMC analyses were performed, using 5 million generations and sampling set adjustment for every 100,000 generations. The remaining parameters were set to default.

LIRs and G-quadruplex prediction. The LIR of each genome was identified in sequence level using Blastn against itself and the identical inverted repeats were collected. For the prediction of G-quadruplexes in *Malassezia*'s mitogenomes, G4Hunter web application, a rewrite of "G4 hunter Python implementation by Jean-Louis Mergny and Amina Bedrat" algorithm, were used setting the window (a) in 25 bp and (b) in 30 bp, and the threshold/G4Hunter score to 1,2[65,66] in both cases.

Data availability

Mitogenomes have been submitted to the NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under accession numbers: ON585844-ON585854.

Received: 16 February 2023; Accepted: 13 April 2023
Published online: 18 April 2023

References

1. Burger, G., Gray, M. W. & Lang, B. F. Mitochondrial genomes: Anything goes. *Trends Genet.* **19**, 709–716 (2003).
2. Hawksworth, D. L. & Lücking, R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr.* **5**, 5–4 (2017).
3. Theelen, B., Christinaki, A. C., Dawson, T. L., Boekhout, T. & Kouvelis, V. N. Comparative analysis of *Malassezia furfur* mitogenomes and the development of a mitochondria-based typing approach. *FEMS Yeast Res.* **21**, foab051 (2021).
4. Kouvelis, V. N. & Hausner, G. Editorial: Mitochondrial genomes and mitochondrion related gene insights to fungal evolution. *Front. Microbiol.* **13**, 897981 (2022).
5. Christinaki, A. C. *et al.* Mitogenomics and mitochondrial gene phylogeny decipher the evolution of *Saccharomycotina* yeasts. *Genome Biol. Evol.* **14**, evac073 (2022).
6. Megarioti, A. H. & Kouvelis, V. N. The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.* **12**, 1337–1354 (2020).
7. Bullerwell, C. E. & Lang, B. F. Fungal evolution: The case of the vanishing mitochondrion. *Curr. Opin. Microbiol.* **8**, 362–369 (2005).
8. Čutová, M. *et al.* Divergent distributions of inverted repeats and G-quadruplex forming sequences in *Saccharomyces cerevisiae*. *Genomics* **112**, 1897–1901 (2020).
9. Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **13**, 1055–1059 (2006).
10. Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10864–10869 (2005).
11. Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.* **6**, e1000861 (2010).
12. Theelen, B. *et al.* Malassezia ecology, pathophysiology, and treatment. *Med. Mycol.* **56**, S10–S25 (2018).
13. Findley, K. *et al.* Topographic diversity of fungal and bacterial communities in human skin. *Nature* **498**, 367–370 (2013).

14. Vijaya Chandra, S. H., Srinivas, R., Dawson, T. L. Jr. & Common, J. E. Cutaneous *Malassezia*: Commensal, pathogen, or protector?. *Front. Cell. Infect. Microbiol.* **10**, 614446 (2020).
15. Gaitanis, G., Magiatis, P., Hantschke, M., Bassukas, I. D. & Velegraki, A. The *Malassezia* genus in skin and systemic diseases. *Clin. Microbiol. Rev.* **25**, 106–141 (2012).
16. Rhimi, W., Theelen, B., Boekhout, T., Otranto, D. & Cafarchia, C. *Malassezia* spp. yeasts of emerging concern in fungemia. *Front. Cell. Infect. Microbiol.* **10**, 370 (2020).
17. Nash, A. K. *et al.* The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* **5**, 153 (2017).
18. Limon, J. J. *et al.* Malassezia is associated with Crohn's disease and exacerbates colitis in mouse models. *Cell Host Microbe* **25**, 377-388.e6 (2019).
19. Gao, R. *et al.* Dysbiosis signature of mycobiota in colon polyp and colorectal cancer. *Eur. J. Clin. Microbiol. Infect. Dis.* **36**, 2457–2468 (2017).
20. Gamal, A. *et al.* The mycobiome: Cancer pathogenesis, diagnosis, and therapy. *Cancers (Basel)* **14**, 2875 (2022).
21. Aykut, B. *et al.* The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* **574**, 264–267 (2019).
22. Abdillah, A. & Ranque, S. Chronic diseases associated with *Malassezia* yeast. *J. Fungi (Basel)* **7**, 855 (2021).
23. Amend, A. From dandruff to deep-sea vents: *Malassezia*-like fungi are ecologically hyper-diverse. *PLoS Pathog.* **10**, e1004277 (2014).
24. Wu, G. *et al.* Genus-wide comparative genomics of *Malassezia* delineates its phylogeny, physiology, and niche adaptation on human skin. *PLoS Genet.* **11**, e1005614 (2015).
25. Xu, J. *et al.* Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18730–18735 (2007).
26. Zhu, Y. *et al.* Proteogenomics produces comprehensive and highly accurate protein-coding gene annotation in a complete genome assembly of *Malassezia sympodialis*. *Nucleic Acids Res.* **45**, 2629–2643 (2017).
27. Lorch, J. M. *et al. Malassezia vespertilionis* sp. nov.: a new cold-tolerant species of yeast isolated from bats. *Persoonia* **41**, 56–70 (2018).
28. Gioti, A. *et al.* Genomic insights into the atopic eczema-associated skin commensal yeast *Malassezia sympodialis*. *MBio* **4**, e00572-12 (2013).
29. Black, B., Lee, C., Horianopoulos, L. C., Jung, W. H. & Kronstad, J. W. Respiring to infect: Emerging links between mitochondria, the electron transport chain, and fungal pathogenesis. *PLoS Pathog.* **17**, e1009661 (2021).
30. Verma, S., Shakya, V. P. S. & Idnurm, A. Exploring and exploiting the connection between mitochondria and the virulence of human pathogenic fungi. *Virulence* **9**, 426–446 (2018).
31. Korovesi, A. G., Ntertilis, M. & Kouvelis, V. N. Mt-rps3 is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. *Mol. Phylogenet. Evol.* **127**, 74–86 (2018).
32. De Chiara, M. *et al.* Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biol.* **18**, 49 (2020).
33. Zhao, X. Q. *et al.* Complete mitochondrial genome of the aluminum-tolerant fungus *Rhodotorula taiwanensis* RS1 and comparative analysis of *Basidiomycota* mitochondrial genomes. *Microbiologyopen* **2**, 308–317 (2013).
34. Aguileta, G. *et al.* High variability of mitochondrial gene order among fungi. *Genome Biol. Evol.* **6**, 451–465 (2014).
35. Moreno-Indias, I. *et al.* Statistical and machine learning techniques in human microbiome studies: Contemporary challenges and solutions. *Front. Microbiol.* **12**, 635781 (2021).
36. Kouvelis, V. N., Ghikas, D. V. & Typas, M. A. 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 (2004).
37. Seif, E. R., Forget, L., Martin, N. C. & Lang, B. F. Mitochondrial RNase P RNAs in ascomycete fungi: Lineage-specific variations in RNA secondary structure. *RNA* **9**, 1073–1083 (2003).
38. Gerhold, J. M., Aun, A., Sedman, T., Jõers, P. & Sedman, J. Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol. Cell* **39**, 851–861 (2010).
39. Liu, X., Wu, X., Tan, H., Xie, B. & Deng, Y. Large inverted repeats identified by intra-specific comparison of mitochondrial genomes provide insights into the evolution of *Agrocybe aegerita*. *Comput. Struct. Biotechnol. J.* **18**, 2424–2437 (2020).
40. Valach, M. *et al.* Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Res.* **39**, 4202–4219 (2011).
41. Locker, J., Rabinowitz, M. & Getz, G. S. Tandem inverted repeats in mitochondrial DNA of petite mutants of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 1366–1370 (1974).
42. Clark-Walker, G. D., McArthur, C. R. & Sriprakash, K. S. Partial duplication of the large ribosomal RNA sequence in an inverted repeat in circular mitochondrial DNA from *Kloeckera africana*. *J. Mol. Biol.* **147**, 399–415 (1981).
43. Nieuwenhuis, M., van de Peppel, L. J. J., Bakker, F. T., Zwaan, B. J. & Aanen, D. K. Enrichment of G4DNA and a large inverted repeat coincide in the mitochondrial genomes of Termitomyces. *Genome Biol. Evol.* **11**, 1857–1869 (2019).
44. Shumard, D. S., Grossman, L. I. & Hudspeth, M. E. Achlya mitochondrial DNA: Gene localization and analysis of inverted repeats. *Mol. Gen. Genet.* **202**, 16–23 (1986).
45. Goldbach, R. W., Bollen-de Boer, J. E., Van Bruggen, E. F. J. & Borst, P. Replication of the linear mitochondrial DNA of Tetrahymena pyriformis. *Biochim. Biophys. Acta* **562**, 400–417 (1979).
46. Ferandon, C., Chatel, S. E. K., Castandet, B., Castroviejo, M. & Barroso, G. The *Agrocybe aegerita* mitochondrial genome contains two inverted repeats of the *nad4* gene arisen by duplication on both sides of a linear plasmid integration site. *Fungal Genet. Biol.* **45**, 292–301 (2008).
47. Pantou, M. P., Kouvelis, V. N. & Typas, M. A. The complete mitochondrial genome of *Fusarium oxysporum*: Insights into fungal mitochondrial evolution. *Gene* **419**, 7–15 (2008).
48. Palmer, J. D. Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet.* **6**, 115–120 (1990).
49. Jansen, R. K. *et al.* Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **395**, 348–384 (2005).
50. Hao, W. From genome variation to molecular mechanisms: What we have learned from yeast mitochondrial genomes?. *Front. Microbiol.* **13**, 806575 (2022).
51. Saccone, C., De Giorgi, C., Gissi, C., Pesole, G. & Reyes, A. Evolutionary genomics in Metazoa: The mitochondrial DNA as a model system. *Gene* **238**, 195–209 (1999).
52. Sahayasheela, V. J., Yu, Z., Hidaka, T., Pandian, G. N. & Sugiyama, H. Mitochondria and G-quadruplex evolution: An intertwined relationship. *Trends Genet.* **39**, 15–30 (2023).
53. Guého-Kellermann, E., Boekhout, T. & Begerow, D. Biodiversity, phylogeny and ultrastructure. in *Malassezia and the Skin* 17–63 (Springer, 2010).
54. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
55. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
56. Jin, J.-J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).

57. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
58. Chan, P. P. & Lowe, T. M. TRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1–14 (2019).
59. Lang, B. F., Laforest, M.-J. & Burger, G. Mitochondrial introns: A critical view. *Trends Genet.* **23**, 119–125 (2007).
60. Kuraku, S., Zmasek, C. M., Nishimura, O. & Katoh, K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* **41**, W22–W28 (2013).
61. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).
62. Wilgenbusch, J. C. & Swofford, D. Inferring evolutionary trees with PAUP. *Curr. Protoc. Bioinform.* **1**, 6–4 (2003).
63. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
64. Abascal, F., Zardoya, R. & Posada, D. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
65. Brázda, V. *et al.* G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* **35**, 3493–3495 (2019).
66. Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**(4), 1746–1759 (2016).

Acknowledgements

The authors wish to thank Prof. Dr. Ferry Hagen for his help with sequencing some of the strains and Dr. Jeffrey M. Lorch for providing isolate CBS 17886. Authors ACC and VNK would like to thank the Federation of European Microbiological Societies (FEMS) which nominated ACC the 'research and training Grant—FEMS-GO-2020–199' for this project. This work was supported by "Special Account for Research Grants" of National and Kapodistrian University of Athens under Research Program (code no. 16673). Part of this work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

Author contributions

A.C.C. and B.T. performed the experiments and wrote the main manuscript text. A.C.C., B.T., and A.Z. performed data implementation. S.C. and J.C. provided necessary data. V.N.K., A.C.C., B.T, and T.B. designed the work and edited the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-33486-4.

Correspondence and requests for materials should be addressed to V.N.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.