FEMS Yeast Research, 21, 2021, foab051

https://doi.org/10.1093/femsyr/foab051
Advance Access Publication Date: 25 September 2021
Research Article

RESEARCH ARTICLE

Comparative analysis of *Malassezia furfur* mitogenomes and the development of a mitochondria-based typing approach

Bart Theelen[1,†], Anastasia C. Christinaki[1,2,†], Thomas L. Dawson, Jr[3,4], Teun Boekhout[1,5] and Vassili N. Kouvelis[2,*,‡]

[1]Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands, [2]Department of Genetics and Biotechnology, Faculty of Biology, National and Kapodistrian University of Athens, Panepistimiopolis, Athens 15701, Greece, [3]Agency for Science, Technology, and Research (A*STAR), Skin Research Institute of Singapore (SRIS), 11 Mandalay Rd, #17-01, Singapore 308232, Singapore, [4]Center for Cell Death, Injury and Regeneration, Departments of Drug Discovery and Biomedical Sciences and Biochemistry and Molecular Biology, Medical University of South Carolina, 280 Calhoun St, Charleston, SC, 29425, USA and [5]Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands

*Corresponding author: National and Kapodistrian University of Athens, Faculty of Biology, Department of Genetics and Biotechnology, Panepistimiopolis, Athens, 15701, Greece. Tel: +302107274488; E-mail: kouvelis@biol.uoa.gr

One sentence summary: A total of 20 *M. furfur* mt genomes show a highly conserved gene order but vary in intron abundance and intergenic regions, revealing promising multicopy target regions for typing.

†These authors have contributed equally to this work.

Editor: Miguel Teixeira

‡Vassili N. Kouvelis, https://orcid.org/0000-0001-6753-0872

ABSTRACT

Malassezia furfur is a yeast species belonging to Malasseziomycetes, Ustilaginomycotina and Basidiomycota that is found on healthy warm-blooded animal skin, but also involved in various skin disorders like seborrheic dermatitis/dandruff and pityriasis versicolor. Moreover, *Malassezia* are associated with bloodstream infections, Crohn's disease and pancreatic carcinoma. Recent advances in *Malassezia* genomics and genetics have focused on the nuclear genome. In this work, we present the *M. furfur* mitochondrial (mt) genetic heterogenicity with full analysis of 14 novel and six available *M. furfur* mt genomes. The mitogenome analysis reveals a mt gene content typical for fungi, including identification of variable mt regions suitable for intra-species discrimination. Three of them, namely the *trnK–atp6* and *cox3–nad3* intergenic regions and intron 2 of the *cob* gene, were selected for primer design to identify strain differences. *Malassezia furfur* strains belonging to known genetic variable clusters, based on AFLP and nuclear loci, were assessed for their mt variation using PCR amplification and sequencing. The results suggest that these mt regions are excellent molecular markers for the typing of

Received: 5 July 2021; **Accepted:** 29 September 2021

M. furfur strains and may provide added value to nuclear regions when assessing evolutionary relationships at the intraspecies level.

Keywords: *Malassezia furfur*; mitochondrial genome; typing; phylogeny

INTRODUCTION

The basidiomycetous yeast genus *Malassezia* is the most dominant fungal component of the human skin microbiome (Findley *et al.* 2020). Their presence on human skin is usually of a commensal nature, but may also cause skin diseases including seborrheic dermatitis (SD)/dandruff, pityriasis versicolor (PV), atopic dermatitis and Malassezia folliculitis (Theelen *et al.* 2018; Saunte, Gaitanis and Hay 2020). Recently, *Malassezia* has attracted increased attention as new insights point towards a more invasive role in other parts of the human body and the genus has been linked to other pathologies, such as Inflammatory Bowel Disease (IBD) and pancreatic cancer (Aykut *et al.* 2019; Limon *et al.* 2019; Spatz and Richard 2020). Although *Malassezia*'s involvement in bloodstream infections (BSIs) of immunocompromised individuals is not a new finding (Brooks and Brown 1987; Kaneko *et al.* 2012) it is of emerging concern, especially for neonates (Iatta *et al.* 2014; Theelen *et al.* 2018). The use of fluconazole as a prophylactic in clinical applications may be a factor contributing to the increased incidence of *Malassezia* BSIs. One major feature of *Malassezia* yeasts is the requirement of lipids for growth, as most of their gene repertoire for carbohydrate metabolism has been lost. As the standard culture media in most clinics do not include lipid supplementation, the role of *Malassezia* in BSIs is likely underestimated (Rhimi *et al.* 2020; Chen *et al.* 2020; Theelen *et al.* 2018).

Malassezia species are not only commensals and pathogens to humans, but also inhabit all warm-blooded animals, with some species found exclusively on animals (Theelen *et al.* 2018; Guillot and Bond 2020). For instance, the most recently described *Malassezia* species, *Malassezia vespertilionis*, was isolated from a bat, illustrating the diverse host-backgrounds of these yeasts (Lorch *et al.* 2018). Based on culture-independent sequencing data, *Malassezia* was found to be present even in diverse habitats such as terrestrial and marine ecosystems including deep-sea sediments, soils, corals, sponges, nematodes and cone snails (Amend 2014; Theelen *et al.* 2018).

Within the genus *Malassezia*, *Malassezia furfur* is found on both human and animal skin and has been implicated in various skin diseases, such as SD and PV (Guého *et al.* 1998; Batra *et al.* 2005; Theelen *et al.* 2018; Saunte, Gaitanis and Hay 2020), and is the most frequently observed causative species (Rhimi *et al.* 2020) in *Malassezia* BSI. Amplified fragment length polymorphism (AFLP) analyses indicated that the species is genetically heterogenous and implies the presence of multiple physiologically diverse strains. Previous studies showed 4–8 genetically diverse clusters which may be linked to their hosts (Theelen *et al.* 2001; Gupta *et al.* 2004), where isolates from deep seated infections primarily belonged to one AFLP-genotype (Theelen *et al.* 2001; Gupta *et al.* 2004).

Although, thus far nuclear-based molecular markers have been employed for isolate identification and typing (Sugita *et al.* 2003, 2005; Gaitanis, Robert and Velegraki 2006; Honnavar *et al.* 2020), the use of mitochondrial (mt) markers may provide a valuable alternative approach. Mitochondria are semi-autonomous organelles that originated from an alpha-proteobacterium as a result of an endosymbiotic event, and they have their own mt genome (Archibald 2015) present in multiple copies per cell. Fungal mt genomes typically contain 14 conserved protein coding genes associated with cellular respiration and ATP production (*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 (*trns*; Kouvelis, Ghikas and Typas 2004). Additionally, a variable gene encoding a ribosomal protein of the small ribosomal subunit (*rps*3) is commonly found in fungi (Korovesi, Ntertilis and Kouvelis 2018). The presence of introns which often contain open reading frames (ORFs) and diverse intergenic regions are mostly responsible for the size variability in mt genomes (De Chiara *et al.* 2020; Megarioti and Kouvelis 2020). Mt variable regions have been used for inter-species discrimination in the fungal kingdom including in *Cryptococcus* (Kortsinoglou *et al.* 2019) and *Lecanicillium* (Kouvelis, Sialakouma and Typas 2008). Moreover, they have been used for intra-species discrimination within *Saccharomyces cerevisiae* (Wolters, Chiu and Fiumera 2015) and *Metarhizium anisopliae* (Ghikas, Kouvelis and Typas 2006). An increasing number of studies have provided insight in *M. furfur* strain fingerprinting with an emphasis on nuclear genetic markers, due to the limited number of available mt genomes (Gupta *et al.* 2004; Gaitanis, Bassukas and Velegraki 2009; Zhang *et al.* 2010). Considering the previously observed genetic intra-species variation in *M. furfur* and proven value of mt target regions for assessing this in other taxa, various mt loci may prove useful for *M. furfur* strain typing based on mt genetic variation.

To define *M. furfur* mt genome variability we performed an analysis of 14 *M. furfur* mt genomes from a Whole Genome Sequencing (WGS) project (unpublished data), including their annotation. Additionally, all available *M. furfur* mt genomes were added (Wu *et al.* 2015) for a complete comparative mt genome analysis. Through this approach, the most variable mt regions suitable for intra-species discrimination were identified. Three of these domains were then used for primer design and amplification. A selected number of strains, representing known genetic variation based on AFLP and nuclear loci (Theelen *et al.* 2001; Gupta *et al.* 2004), were assessed for their variation in these mt variable regions, and evaluated for their usefulness for the typing of *M. furfur* strains of clinical and other origins.

MATERIALS AND METHODS

Malassezia furfur strains

In this study, 20 *M. furfur* strains were used for the *in silico* mt genome analysis. In addition to the six *M. furfur* published mt genomes (Wu *et al.* 2015), 14 mt genomes were retrieved from unpublished WGS data (Table 1). A total of 23 other *M. furfur* strains were included in the analysis to assess the intraspecific variation (Table 1).

Culture and DNA extraction

Malassezia furfur cells were grown on modified Dixon agar (mDA; Guého, Boekhout and Begerow 2010) at 30°C for 48 h. For Illumina sequencing purposes, cells were harvested in 50 mL tubes. DNA was extracted with the QIAGEN Genomic DNA Purification procedure for yeast samples (Qiagen, Hilden, Germany), with some modifications. In detail, lyticase incubation was performed for 2 h at 30°C, and RNase/Proteinase incubation followed for 2 h at 55°C. Genomic DNA was then purified with

Table 1. *Malassezia furfur* strains used for the *in silico* mitogenomic comparative analysis and for *in vitro* experiments, their source and location. Accession numbers of *M. furfur* known and newly acquired, mt genomes (Wu *et al.* 2015; this study) and their mt genome sizes are also presented. Asterisk (*) denotes that these strains have been used only for the whole mt genome comparative analysis. NA: Non-Available, PV: pityriasis versicolor and SD: seborrheic dermatitis. A hashtag (#) indicates strains originating from deep-seated human body parts.

Strain	Source	Country	Mt genome accession number	Mt genome size (bp)	Current study
CD864	Chronic pruritic skin disease, poodle	Brazil	MW683316	48 849	*In silico*/PCR
MAL18#	Blood	Italy	MW683317	48 977	*In silico*/PCR
MAL24	Arm skin	Italy	MW683318	48 977	*In silico*/PCR
MAL26#	Blood	Italy	MW683319	48 958	*In silico*/PCR
MAL32#	Urine	Italy	MW683320	48 281	*In silico*/PCR
CBS9374*	Chest, healthy human	Canada	MW683314	47 717	*In silico*
CBS9365	Elephant in zoo	France	MW683313	48 188	*In silico*/PCR
CBS4169	Eye lid, man	Netherlands	MW683308	45 715	*In silico*/PCR
CBS6000	Dandruff, man	India	MW683310	49 317	*In silico*/PCR
CBS6001	PV, man	India	MW683311	49 274	*In silico*/PCR
CBS5334I	Infected skin, man	Canada	MW683309	49 217	*In silico*/PCR
CBS1878	Dandruff, man	unknown	KY911081.1	48 161	*In silico*/PCR
CBS4172	Skin, eland	NA	KY911082.1	48 279	*In silico*/PCR
CBS7019	PV, trunk,15-year-old girl	Finland	KY911083.1	49 305	*In silico*/PCR
CBS7710*	Skin, man	Netherlands	KY911084.1	47 901	*In silico*
CBS7982	Skin of ear, healthy man	France	KY911085.1	47 903	*In silico*/PCR
CBS14141#	Catheter, blood, human	France	KY911086.1	48 933	*In silico*/PCR
CBS8735#	Bronchial wash, man	Canada	MW683312	49 046	*In silico*/PCR
PM315	Anal swab, neonate	Germany	MW683321	49 213	*In silico*/PCR
CBS14139#	Urine	France	MW683315	49 242	*In silico*/PCR
CBS9370	Back of healthy individual	Canada	NA	NA	PCR
PM312#	Urine, neonate	Germany	NA	NA	PCR
CBS14140#	Catheter, blood	France	NA	NA	PCR
CBS7854	Seborhoeic scalp, man	Finland	NA	NA	PCR
CBS4162	Ear, pig	NA	NA	NA	PCR
CBS5101	PV, skin scales, man	USA	NA	NA	PCR
CBS7969	Back, Asian elephant	France	NA	NA	PCR
CBS4171	Ear, cow	NA	NA	NA	PCR
CBS4170	Ear, horse	NA	NA	NA	PCR
CBS6046	PV	USA	NA	NA	PCR
CBS6093	Oval-cell variant of CBS5634	NA	NA	NA	PCR
CBS6094	Normal skin of rump	USA	NA	NA	PCR
CBS7043	24-year-old, man	Belgium	NA	NA	PCR
CBS7981	PV, skin, woman	France	NA	NA	PCR
CBS8736	PV lesion, trunk, woman	Canada	NA	NA	PCR
CBS5332	Infected skin, man	Canada	NA	NA	PCR
CBS9369	Scalp, man	Canada	NA	NA	PCR
CBS9585	SD, submammary fold, man	Greece	NA	NA	PCR
CBS7985	Wing, Struthio camelus (ostrich)	France	NA	NA	PCR
CBS9574	PV, back, man	Greece	NA	NA	PCR
CBS9575	SD, back, man	Greece	NA	NA	PCR
CBS9589	SD, nasolabial folds, man	Greece	NA	NA	PCR
CBS9595	SD, back, man	Greece	NA	NA	PCR

Genomic-tip 100/G prep columns according to the manufacturer's instructions. For PCR and sequencing of nuclear IGS and mt loci of all other *M. furfur* strains, DNA extraction was performed with the CTAB method as described by O'Donnell *et al.* (1997) with some modifications: two 10 μL loops of fresh yeast cells were suspended in 750 μL CTAB buffer and mechanically disrupted in a TissueLyser II (Qiagen®) at 30 Hz for 8 min. The mixture was heated for 1 h at 65°C with vigorously agitation (vortex) halfway. After lysis, two purification steps were performed with phenol–chloroform and chloroform, respectively. Extracted DNA was dissolved in 100 μL TE and 10x diluted in MQ water, prior to PCR.

WGS

Illumina reads were generated using the TruSeq v3 PE Cluster Generation Kit and SBS kits, on the Illumina HiSeq2000 system (HCS v2.2.58, RTA v1.18.64) (Illumina, San Diego, CA). Paired end reads were trimmed with Trimomatic v0.33 (Bolger, Lohse and Usadel 2014) and contigs were generated with SPAdes v3.13.0 (Bankevich *et al.* 2012), using standard settings. PacBio reads were generated on the PacBio RSII system. All acquired PacBio reads were further processed and initial contigs were created with the HGAP3 software pipeline of the PacBio SMRT portal, SMRT analysis v3.1 (PACBIO, Menlo Park, CA), with standard

settings. The obtained contigs were further processed using program SeqMan of Lasergene Suite 11 (DNASTAR Inc. Madison, WI; Burland 2000), using the publicly known mt genome of *M. furfur* CBS 1878 KY911081.1 as a reference template to create mt contigs. These were further analysed as described below.

Mt genome and gene order (synteny) analyses

The previously published mt genomes with GenBank accession numbers KY911081.1, KY911082.1, KY911083.1, KY911084.1, KY911085.1 and KY911086.1 were available in annotated versions. The additional 14 mt genomes were annotated in this study and deposited to GenBank (accession numbers: MW683308–MW683321). The mt genomes were retrieved manually by aligning all contigs from the WGS data against the above known mt genomes using tBLASTn/BLASTx/BLASTn (Altschul *et al.* 1990). 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 (Chan and Lowe 2019). The mt circular map of strain CBS5334 was created using the Geneious Prime 2021 (Biomatters, Auckland, New Zealand). A comparative genome analysis was performed to locate the similarities and the differences in mt gene order (synteny) among the examined strains. The 20 mt *M. furfur* genomes were aligned by multiple sequence alignment program MAFFT using the E-INS-i method (Kuraku *et al.* 2013; Katoh, Rozewicki and Yamada 2019), and the mito-*M. furfur* matrix was produced.

Primer design, PCR amplification and sequencing

Based on the mt comparative analysis of the *M. furfur* strains (mito-*M. furfur* matrix), three highly variable regions were identified, i.e. the *trnK–atp6* intergenic region, the *cox3–nad3* intergenic region and the second intron (region between exon2 and exon3) of *cob*, and further used for typing of all strains (Table 1). Therefore, three primer sets were designed, using the program PrimerSelect of Lasergene Suite 11 (DNASTAR Inc. Madison, WI; Burland 2000), and the online bioinformatics software Primer 3 plus, using default settings (Untergasser *et al.* 2012; Table. 2). Additionally, the nuclear ribosomal intergenic spacer (IGS) was added in the analysis using previously published primers as this region is also informative for the typing of *Malassezia* species (Sugita *et al.* 2002, 2003, 2005). The above mentioned mt regions and the IGS region were amplified for the 41 studied strains. PCR amplification reactions were performed with MyFi DNA polymerase (Bioline Meridian BIOSCIENCE, Cincinnati, OH) in a Sensoquest LabCycler according to the manufacturer's instructions. The thermocycling protocol was 5 min at 95°C, 35 cycles of 15 s at 95°C, 20 s at 46°C and 20 s at 72°C for *trnK–atp6* and *cox3–nad3* PCR reactions. For the cob (exon2–exon3) intronic and IGS regions the same protocol was used, but with different annealing temperatures (50°C and 55°C, respectively).

All PCR amplicons were purified with magnetic beads in a Microlab STAR liquid handling System (Hamilton, Reno, NV), and were sequenced in both directions using the BigDye Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific, Waltham, MA) in a 3730xl DNA Analyzer (Applied Biosystems, Waltham, MA). The forward and reversed sequences were analyzed using SeqMan of the Lasergene Suite 11 software package (DNASTAR Inc. Madison, WI; Burland 2000), and similarity searches in the retrieved consensus sequence were performed with BLAST (Altschul *et al.* 1990). The sequences of all amplicons are deposited into GenBank. (Accession numbers for *trnK-atp6* intergenic region: MZ494193–MZ494233; for *cox3–nad3* intergenic region: MZ494234–MZ494274; for *cob* exon2–exon3 [cobi2] intronic region: MZ494275–MZ494315 and for IGS: MZ494316–MZ494356).

Phylogenetic analyses

Nucleotide sequences of all amplified regions were collected to create matrices for phylogenetic trees. The sequences were aligned by Lasergene's MegAlign v.11 program (Burland 2000) using the ClustalV method with default settings. When necessary, alignments were edited manually. Single region matrices (IGS and the three selected mt regions) and the concatenated dataset of mt regions were created for phylogenetic purposes. The phylogenetic tree constructions were performed using Neighbor Joining (NJ) and Bayesian Inference (BI) methods through PAUP4 (Wilgenbusch and Swofford 2003) and MrBayes (Ronquist and Huelsenbeck 2003) 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 JmodelTest (ver. 2.0; Guindon and Gascuel 2003; Darriba *et al.* 2012). The Bayesian Information Criterion (BIC) was applied, and the best nucleotide substitution model was found. In detail, HKY + G and TPM3uf + I + G were applied for the individual datasets and for the concatenated dataset, respectively. In all cases, 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 in all cases. Additionally, all mt coding genes for the 20 strains of *M. furfur* with known complete mitogenomes were aligned as a single matrix, and NJ and BI methods were employed for the construction of the phylogenetic tree as described in Fig. 5.

RESULTS

Mt comparative analysis

Comparative mt genome analysis revealed a highly conserved gene order among *M. furfur* strains (Table S1, Supporting Information). The gene content for all examined mt genomes consisted of 15 protein coding genes, two ribosomal (rRNA) genes and 24 tRNA genes (Fig. 1A). However, the sizes of the mt genomes ranged from 45 715 (strain CBS4169) to 49 317 bp (strain CBS6000; Table 1) attributable to differences in intron abundance and variable intergenic regions.

In detail, introns were located only in *cox1*, *cob* and *rnl* genes. Introns of *cob* and *rnl* were without ORFs, while *cox1* introns contained ORFs. Specifically, the *cox1* gene consisted of five exons. In all examined genomes of this study cox1i1 (*cox1* intron 1) and cox1i4 contained a single ORF, and cox1i3 contained two ORFs. In addition, cox1i2 was ORF-less, but in nine strains, i.e. CBS14141, CBS6001, CBS5334, CBS6000, CD864, MAL18, MAL24, MAL26 and MAL32, an ORF was present. All ORFs encoded putative homing endonucleases (HEs) belonging to the LAGLIDADG family (as defined by Belfort *et al.* 2002). Furthermore, strain CBS14141 lacked cobi2 while all other strains retained it. Finally, in almost all examined strains *rnl* contained three introns of either subtypes IB or IA. In detail, rnli1 was group IB, while the

Table 2. Primers used in this work, the region which they amplify and their reference in literature.

Region amplified	Primer's name	Sequence (5'-3')	Reference
trnK—atp6 intergenic region	MF1trnKF	AACCCATTGAAAAGGAGAAC	Current work
	MF1atp6R	CAAGAGGAGAATGAATAAAGTAAG	
cox3–nad3 intergenic region	MF2cox3F	CTTTTTGAACTTTGAATGTATGGTC	Current work
	MF2nad3R	ATTATGGTTTCTGGATTATTATGGTA	
cob exon2–exon3 intronic region	MF3cob_2F	CTATATGGCGATGCGAGTG	Current work
	MF3cob_3R	AGAGCTGCTAGAATGAATGGTA	
Ribosomal Intergenic Spacer (IGS)	26SF	ATCCTTTGCAGACGACTTGA	Sugita *et al.* (2002)
	5SR	AGCTTGACTTCGCAGATCGG	

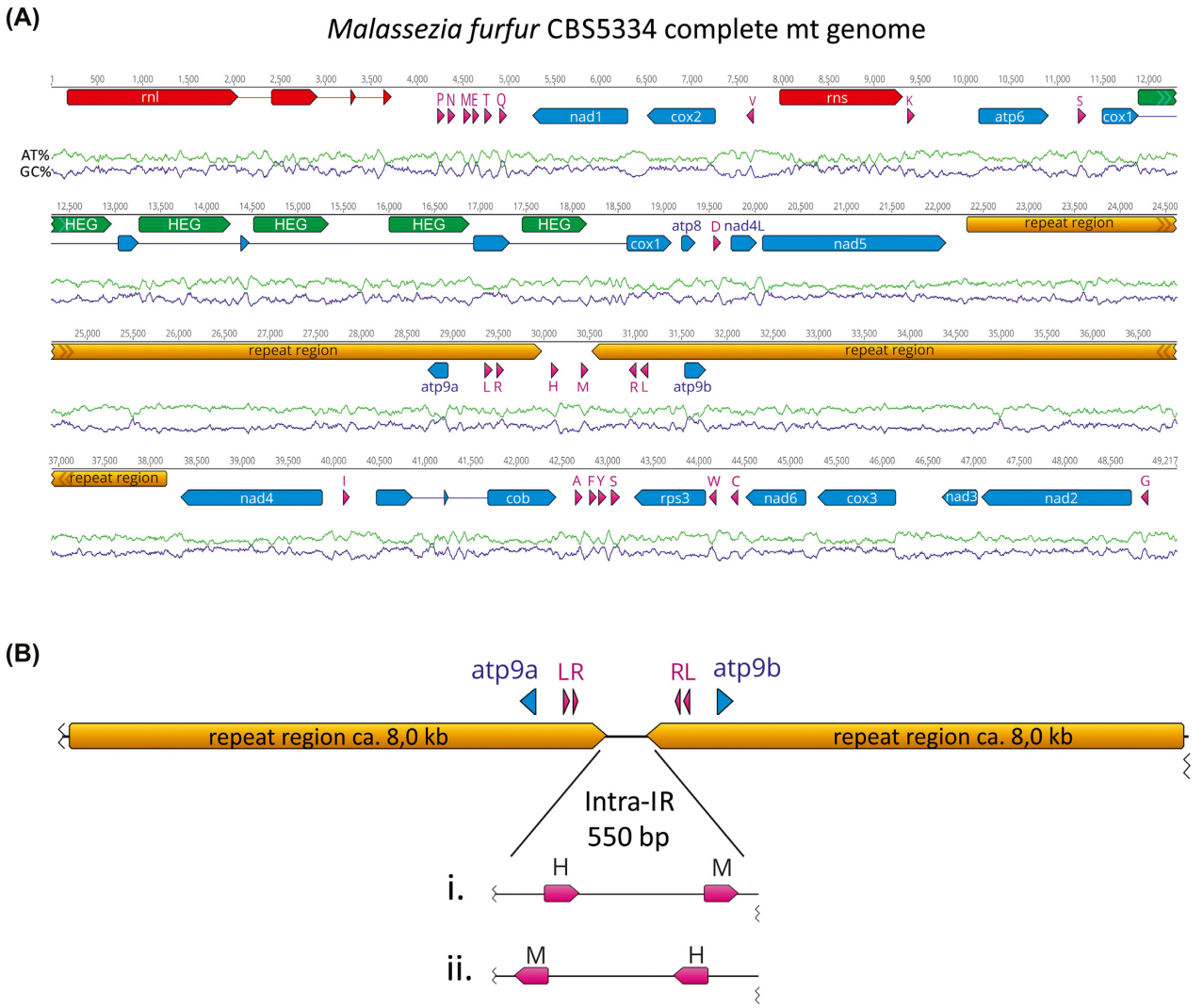

Figure 1. (A) The linear representation of the circular mt genome of *M. furfur* strain CBS5334, beginning with the *rnl* gene. Features are presented by bars and their orientation is indicated by the pointed end. In detail, conserved protein coding genes, rRNA genes, tRNA genes and the inverted repeat are presented in blue, red, pink and orange, respectively. The single amino acid letters represent the respective tRNA genes (e.g. K →*trn*K). Genes *rnl*, *cox*1 and *cob* are interrupted by intronic regions. Introns of the *cox*1 gene contain ORFs encoding putative homing endonucleases (HEs) of LAGLIDADG family. GC and AT content are shown with dark blue and light green colors, respectively. The other mt genomes examined present the same gene order (synteny) but differ in intronic and intergenic region abundance and size, as well as in the intra-IR orientation. **(B)** The large inverted repeat (IR) of the *M. furfur* mt genome (orange), containing the duplicated mt genes *atp*9, *trn*L and *trn*R. The 550 bp intra-IR fragment can be found in both orientations (i/ii) among *M. furfur* strains (Table S1, Supporting Information).

other two, i.e. rnli2 and rnli3, were of group IA. The only exceptions were strains CBS4169 (with only the third intron—rnli3), CBS9374 (containing only rnli1 and rnli2) and CBS9365 (hosting rnli2 and rnli3).

All *M. furfur* strains contained a remarkable feature in their mt genome, namely an approximately 8000 bp region duplicated in inverted orientations and split by a ca. 550 bp fragment. This small fragment was found in two orientations among the different examined strains (Fig. 1B and Table S1, Supporting Information). The inverted repeated (IR) region contained three genes (*atp*9, *trn*L and *trn*R), which were also duplicated. The inta-IR fragment, regardless of its orientation, contained the tRNA genes *trn*M and *trn*H (Fig. 1B).

Based on the aligned mt genomes (mito-*M. furfur* matrix), variability for each of the protein coding genes was assessed and aminoacid sequence divergence was ranging from 0.5% (at gene *nad*3) to 1.6% (at gene *rps*3; Table S2, Supporting Information). However, higher variability (see below) was observed in all intergenic regions, showing better potential for typing. Therefore, the most variable intergenic regions were further analyzed in the following sections.

Strain typing

Following the comparative mt genome analysis, intraspecific variation of 41 *M. furfur* strains was assessed, using the ribosomal IGS region as a nuclear benchmark to compare with three identified mt target regions (see Materials and methods).

Ribosomal IGS

The IGS region was amplified for the selected strains using primers 26SF and 5SR (Table 2). IGS Amplicons contained noticeable size variation ranging from 433 to 491 bp (Table S3, Supporting Information). A phylogenetic tree based on these sequences clustered *M. furfur* strains in seven well supported groups (A1, A2, B, D, E, G and H; Fig. 2). Groups A1 and A2 did not present any significant differentiation, other than one or two Single Nucleotide Polymorphisms (SNPs; 99.8% identity; Table S4, Supporting Information). In closely related groups, such as group A1 and B, IGS amplicons differ in one SNP and in a three nucleotide (nt) deletion in sequences of group B (98.2% identity). On the contrary, IGS sequences of groups A1 and E presented only 81.1% identity due to SNPs and deletions. Interestingly, variation within IGS groups was low and in most cases the sequences were identical. IGS amplicons from group H had distinct sequences compared to sequences of the other *M. furfur* strains (identity ranging from 49.6 to 50.8%), and clustered basal to all other *M. furfur* groups. Additionally, group H was divided into two smaller sub-clusters of three and five strains, respectively, which showed a differentiation of ca. 2.7% divergence in the IGS region. Only strain CBS7985 of cluster H presented a larger divergence which ranged from 5.9 to 4.3% compared to sub-clusters A and B, respectively (Table S4, Supporting Information). For example, IGS sequences of strains CBS9575 (491 bp) and CBS7985 (455 bp) shared only 94.1% identity because of four deletions and two insertions in CBS7985 (Table S4, Supporting Information).

Moreover, with consideration of host background it became evident that IGS group A (A1/A2) contained strains of different sources, including all strains from deep-seated body parts (except MAL18 which was situated at group G). IGS group B consisted of five strains isolated from animals and two associated with human skin disease. Group E exclusively contained strains isolated from diseased skin, although the sample size was low ($n = 5$) and the background of one strain was unknown.

Based on the comparative analysis of the available mitogenomes, the three most promising variable regions were selected for the development of PCR-assays to evaluate their suitability for strain typing in *M. furfur*.

trnK-atp6 intergenic region

trnK–atp6 intergenic region amplicons revealed considerable size variation, from 694 to 777 bp in CBS7982 and CBS6001, respectively, with a mean size of 759 bp (Table S3, Supporting Information). An exception is strain CBS9370 with an amplicon size of only 222 bp resulting from a 580 bp deletion (position 120–700 in the alignment). Among the 41 strains in this study, 31% of the matrix's sites were diverse. *Malassezia furfur* strains grouped in seven well-supported clusters (a–g; Fig. 3A). Cluster c consisted of 20 *M. furfur* strains but without statistical support (Bootstrap values $< 50\%$). Sequence divergence between strains of different clusters ranged from 0.1 to 8.0% (Table S4, Supporting Information). Within observed clusters differentiation among strains was also observed. Cluster g was the most diverse with 13 SNPs. Clusters a and b were exceptions as amplicons differ in only one or two SNPs. Consequently, the mt *trnK–atp6* intergenic region showed an intriguing divergence, which however did not result in a clear clustering pattern.

cox3–nad3 intergenic region

This was the most variable region, with more than 45% of the aligned sequence sites indicating divergence across all strains examined. This divergence may be attributed to amplicon size variability. In detail, the amplicon size ranged from 755 (strain CBS9595) to 867 bp (strain CBS14141), due to deletions and insertions (Table S3, Supporting Information). As a result, each strain provided a unique sequence for the *cox*3–*nad*3 intergenic region. The only exceptions were strains CBS4169 and CBS4170, which were identical (100%—Table S4, Supporting Information). Phylogenetic analyses based on the matrix of this region classified strains in seven well-separated clades (clusters h–n; Fig. 3B). The five strains belonging to cluster n showed the highest intra- and inter- specific divergence (only 97.5% and ca. 78.1% identity, respectively—Table S4, Supporting Information).

cobi2 region

Primers MF3cob_2F and MFcob_3R were used to clarify the intron presence in the *cob* gene. Amplification resulted in two types of products: one of smaller size (375–387 bp) without an intron, and a second larger-sized type (801–826 bp; Table S3, Supporting Information) with a 445 bp ORF-less ID intron inserted into position 429 of the *cob* gene. In the exon sequence of the amplified region (end of exon2–beginning of exon3), more than 80 variable sites were detected. The complete amplified region was used in the phylogenetic tree construction (Fig. 3C). The 41 *M. furfur* strains grouped in six clusters (o–t). The strains that did not contain an intron in that region were located basally in the phylogenetic tree (clusters o and t). Interestingly, cluster o consisted of strains with both types of the amplified region, i.e. strains containing the ID intron (PM312 and PM315) and intronless ones (CBS5332, CBS7854, CBS14140 and CBS14141).

Concatenated mt phylogenies

In order to assess how well each of the chosen individual mt regions reflect genetic variation and clustering between *M. furfur* strains for typing purposes when compared with a broader multi-locus mt approach, the phylogeny of a concatenated mt dataset was constructed, based on the combined number of

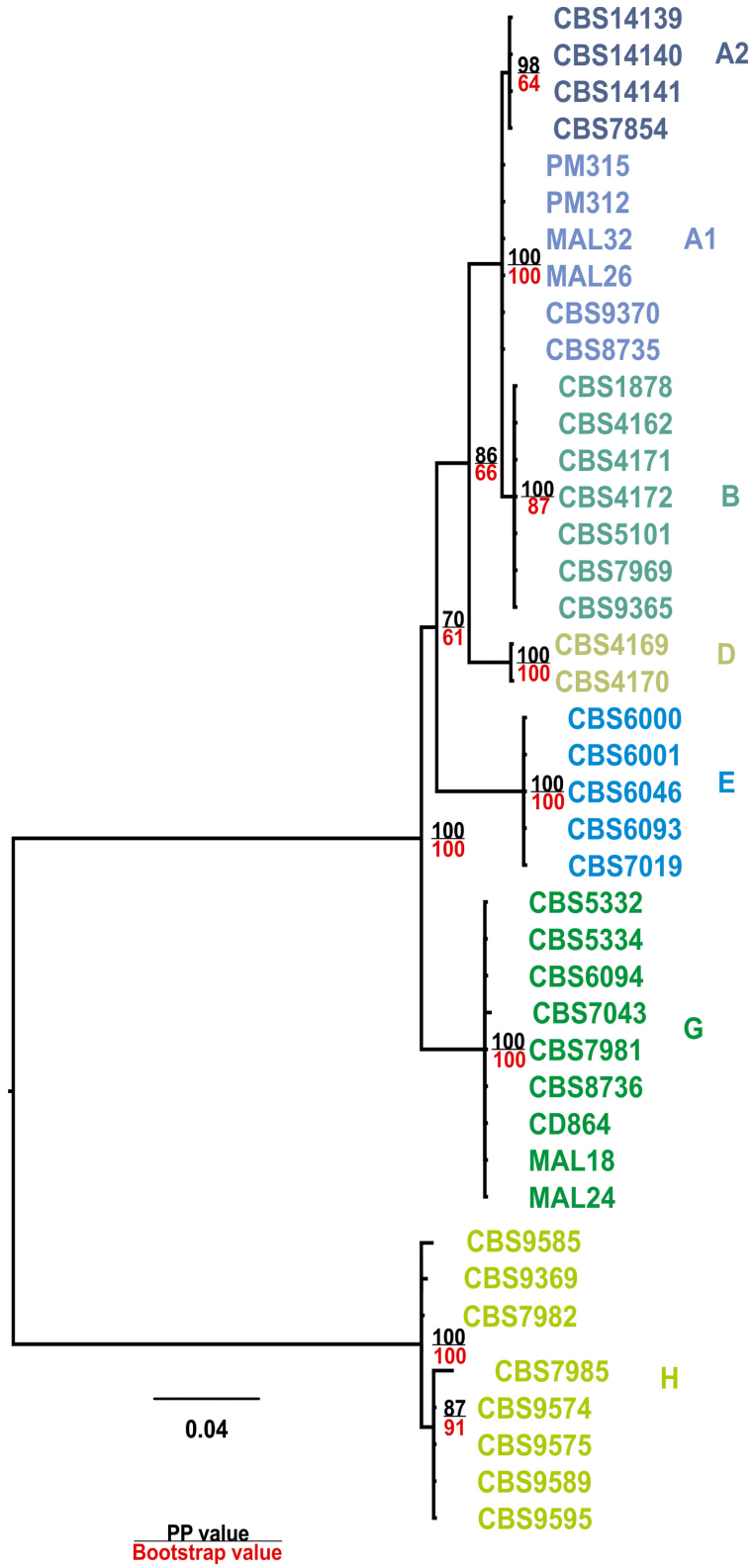

Figure 2. Phylogenetic tree of 41 *M. furfur* strains as produced by BI based on IGS. The NJ tree is identical. Phylogenetic relationships among taxa are well-supported. Posterior probabilities (PP) and Bootstrap support (BS) are presented in black and red numbers, respectively. IGS groups (A, B, D, E, G and H) are marked with different colors.

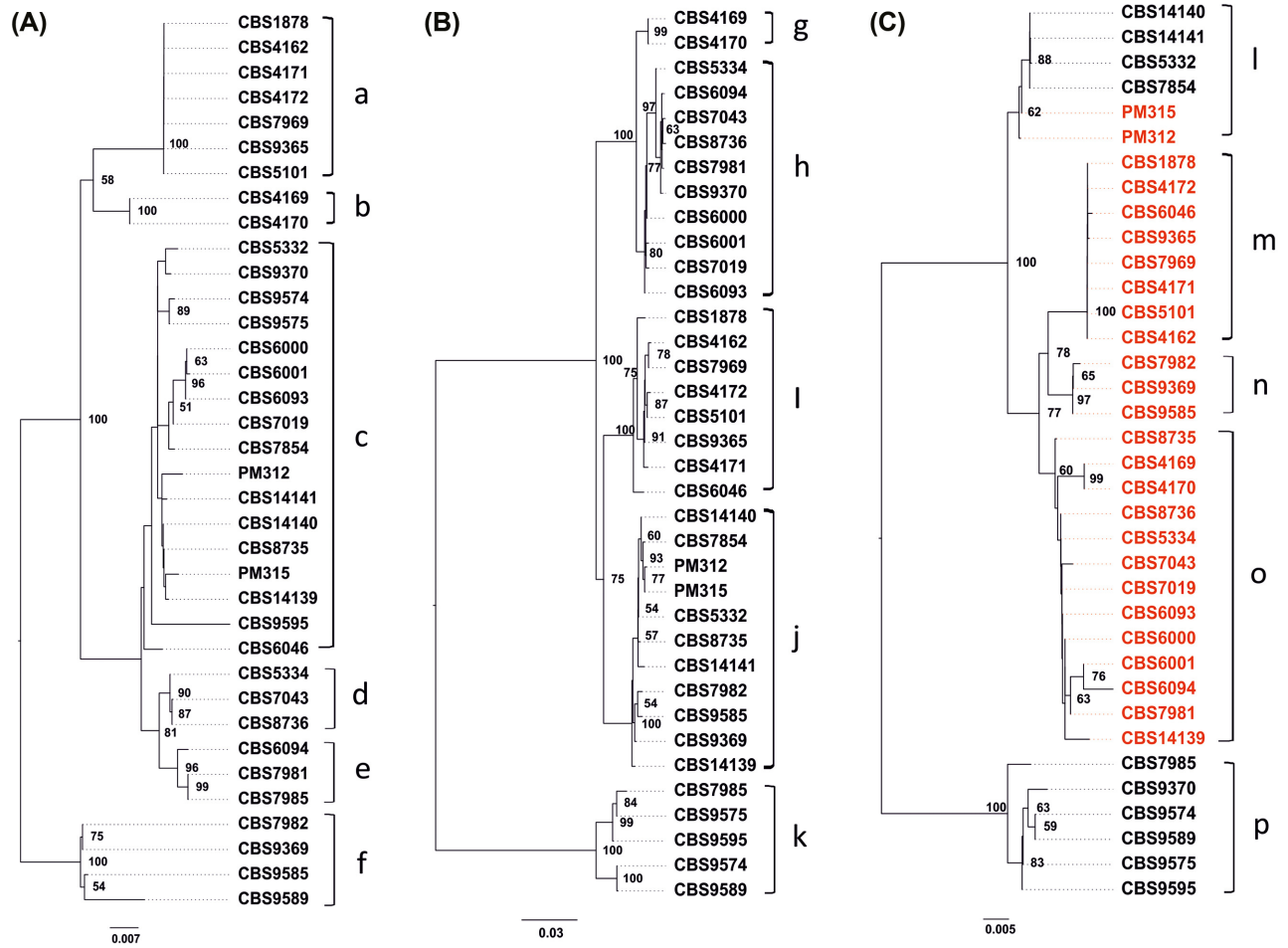

Figure 3. Phylogenetic trees as produced by BI analysis (NJ trees were identical) based on **(A)** *trnK–atp6* mt intergenic region, **(B)** *cox3–nad3* mt intergenic region and **(C)** *cob* exon2–exon3 mt intronic region. Strains containing an intron within these amplicons are shown in red. Node credibility is presented with Bootstrap values with numbers (Bootstrap values > 50%).

informative sites of all three individual mt regions. The concatenated matrix based phylogenetic tree was well-supported (Fig. 4), and mt-based relationships of the strains well-defined. *Malassezia furfur* strains were classified in five main groups and 10 subgroups (Fig. 4). Strains belonging to group I were significantly divergent using the concatenated dataset, although the clustering received high support (100% PP and 96% BS). An exception is CBS6046, which may be considered as another subgroup basal to the group I. Group II consisted of four subgroups (II1, II2, II3 and II4) and group III of three (III1, III2 and III3). Group II was a sister clade to group I (with support of 100% and 97% PP and BS, respectively), while group III was a sister clade to both (100% support for both statistical methodologies applied). Strain CBS9370 was located basally to the previous three groups and was designated as group IV. Strains belonging to group V were the most diverse.

If the origin of the strains was taken into account, it became evident that group I consisted of five strains isolated from animal and two from human skin (the source of the basal strain CBS6046 was from diseased human skin). Strains found on human skin could be also found in subgroups II1, II2, II4 and III2. Strains originating from deep-seated body sites were all positioned in subgroups II3, III1 and III3, though these clusters were not exclusive for these strain backgrounds. The strains originating from diseased skin of human individuals from Greece comprised group V with the addition of a strain from an ostrich located in France.

Based on the available 20 mt genomes of *M. furfur* strains, phylogenetic relationships were also determined using the 15 concatenated protein coding genes, in order to evaluate phylogenetic differences between a protein coding gene-based approach (Fig. 5) and the concatenated intergenic target regions (Fig. 4). However, the limited number of available mt genome sequences per genotype, prevents any assumptions for the relationships between genotype and host background.

DISCUSSION

Detailed knowledge on *Malassezia* mt genomes and genes is scarce. A study focusing on *M. sympodialis*, explored the mt genome of this species (Gioti *et al.* 2013) revealing a typical mt genome of 38 622 bp containing 15 protein-coding genes, two rRNA genes and 25 tRNAs (Gioti *et al.* 2013). Another study on comparative genomics of various *Malassezia* species compared general phylogenetic topologies between nuclear and mt genes and concluded that both trees corresponded well at an inter-species level. However, differences were observed for some species, including *M. furfur*, at the intra-species level, albeit based on only very few strains (Wu *et al.* 2015). In our comparative analysis of 20 mitogenomes it is shown that the *M.*

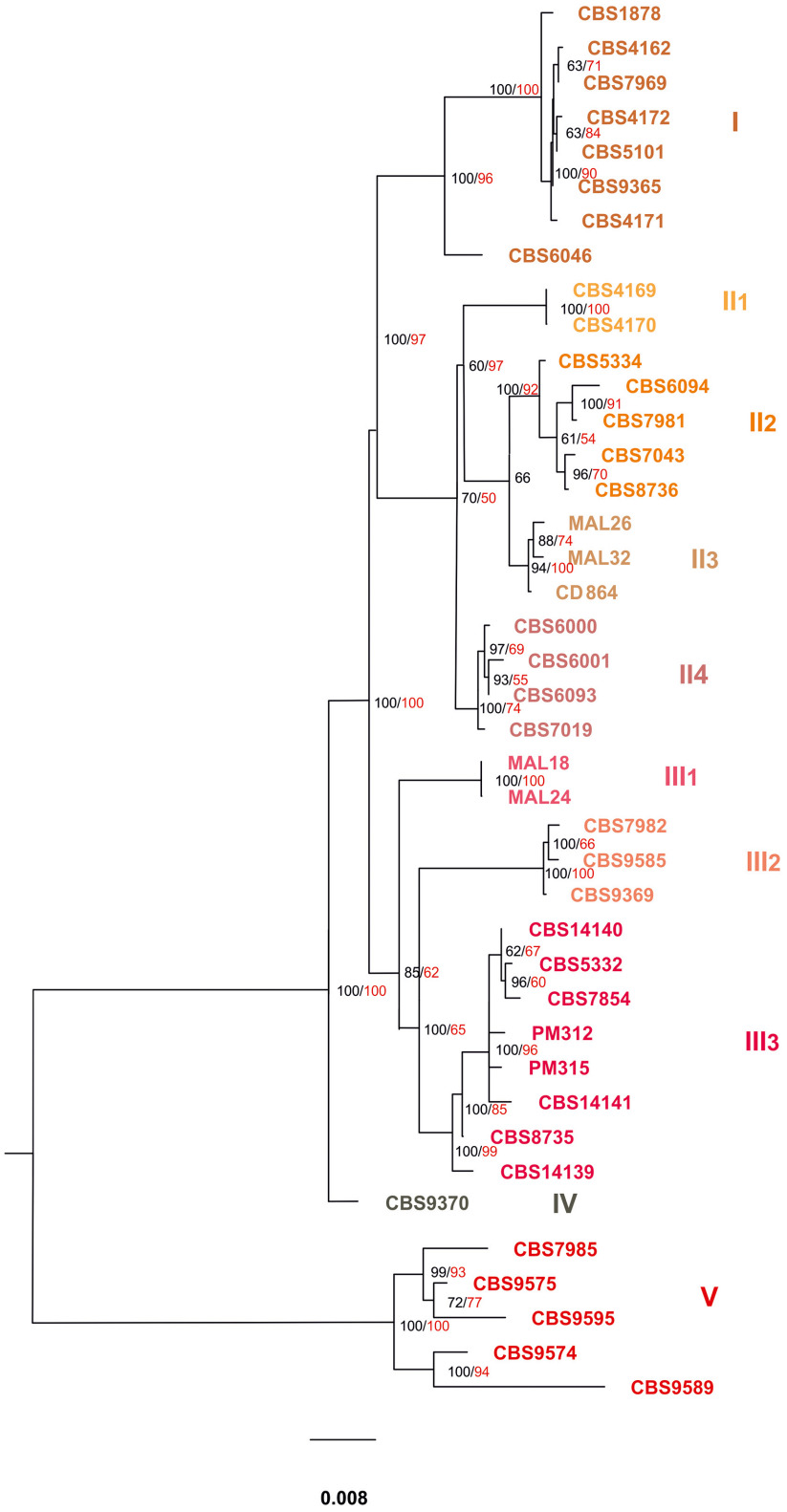

Figure 4. Phylogenetic tree of *M. furfur* strains as produced by BI based on the concatenated dataset of *trnK–atp6* and *cox3–nad3* mt intergenic regions, and *cob* exon2–exon3 mt intronic region. The corresponding tree using NJ analysis is identical. Phylogenetic relationships among taxa are well supported with high PP (black numbers) and BS (red numbers) values. Mt groups and subgroups (I–V) are distinguished by color.

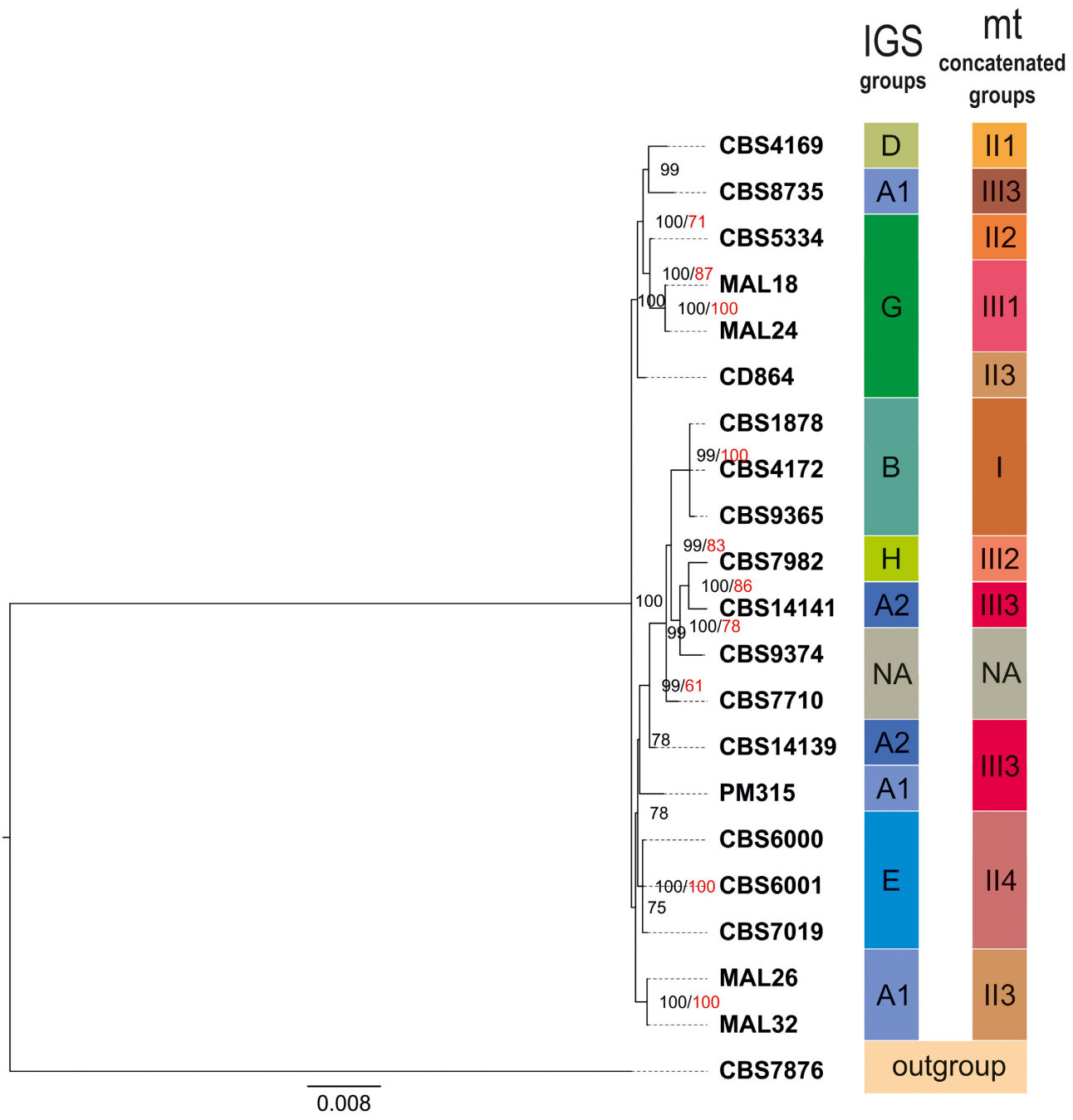

Figure 5. Phylogenetic tree of the 20 *M. furfur* strains with known mt genome as produced by BI based on the 15 mt genes. The species *M. obtusa* CBS7876 (GenBank assembly accession: GCA_001264985.1) was used as outgroup. For the NJ method, reliability of nodes was evaluated using 10,000 bootstrap iterations. For BI analysis, the evolutionary model, as found using the program was GTR + I + G. A total of four independent MCMCMC analyses were performed, using 5 million generations and sampling set adjustment for every 100 000 generations. Clade support (Posterior Probabilities) is shown (black numbers). Whenever an NJ identical topology exists, Bootstrap Support values (red numbers) are presented. Next to the phylogenetic tree, the two bars represent the corresponding IGS (Fig. 2) and concatenated mt (Fig. 4) phylogenetic groups' coloring. NA: non-available

furfur mt genome has a conserved gene content and order but is highly variable in size, ranging from 45 715 to 49 317 bp. This diversity has also been observed in other fungal species such as *Rhizophagus irregularis*, *Cordyceps militaris* and *Lachancea thermotolerans* (Formey *et al.* 2012; Freel, Friedrich and Schacherer 2015; Zhang *et al.* 2017), with a size divergence of approximately 3 kb. The largest size variation has been detected in the mt genome of *S. cerevisiae*, with a size variation of approximately 22 kb among strains (Foury *et al.* 1998; De Chiara *et al.* 2020). In other studies, correlation has been shown between genome size fluctuation and both intron abundance and size diversity of intergenic regions, in accordance with this study (Wolters, Chiu and Fiumera 2015; Deng *et al.* 2018; De Chiara *et al.* 2020). From our study, it appears that strains with larger mt genomes belong to groups II3, II4, III1 and III3 of the concatenated mt phylogenetic tree and groups A1, A2, E and G of IGS tree. Thus, it is evident that mt genome expansion and reduction in size is unlikely to be a unidirectional evolutionary process.

We observed a duplicated and inverted region of ca. 8000 bp with three included genes (*atp*9, *trn*L and *trn*R), interrupted by a ca 550 bp intra-IR fragment, found in two orientations among different strains. Similarly, an inverted repeat region was previously also identified for *M. sympodialis*, with a size of 5900 bp (Gioti *et al.* 2013; Zhu *et al.* 2017). These large IRs are common in fungal mt genomes and have been described for a plethora of fungal genera, including the ascomycetous yeast genus *Candida*, significantly fluctuating in length from 109 bp in *Candida salmanticensis* to 14 379 bp in *Candida maltosa*, with varying gene content, and having a substantial impact on mt genome size (Valach *et al.* 2011). They have also been linked to conversions between circular and linear mt genome forms (Valach *et al.* 2011). Large inverted repeats have also been described for other fungi such as the genus *Termitomyces*, where the IR was half the size of the complete mt genome, duplicating several genes (Nieuwenhuis *et al.* 2019). Another example is the mushroom *Agrocybe aegerita*, where the presence of the IRs was hypothesized to mediate intramolecular recombination (Liu *et al.* 2020). Biological implications of large IRs need further examination and requires a genus wide approach in *Malassezia*.

In addition to previous observations (Theelen *et al.* 2001; Gupta *et al.* 2004), an ongoing exploration of the sub-species variation in *M. furfur* is further unraveling the genetic heterogeneity of this species based on nuclear loci. The IGS region highlights this variability and was used in this study as a nuclear reference locus for the known nuclear intraspecific variation. The multicopy nature of mt genomes, as well as known differences in evolutionary rates, make them promising targets for developing molecular markers for typing approaches (Ghikas, Kouvelis and Typas 2010; Kortsinoglou *et al.* 2019, 2020). Based on the comparative analysis of the analyzed 20 *M. furfur* mt genomes, three highly variable regions of interest were discovered and were assessed for a total of 41 strains, representing intra-species variation, which was compared to the IGS based clustering. Firstly, the *trnK-atp6* mt intergenic region is more diverse than the IGS region (Table S4, Supporting Information), and therefore, has potential for genotyping. Moreover, the tree topology is different when compared to the respective tree of the mt concatenated matrix (Figs 3A and 4). In this amplified domain a number of nucleotide substitutions were detected (Table S4, Supporting Information) but this differentiation may reduce applicability for assessing intra-species diversity of the *M. furfur* mt genome as it provides branches that are not well supported (Fig. 3A). Possible usefulness of this region for typing should be further examined with a larger number of strains. Secondly, the *cox*3–*nad*3 region demonstrated the best strain discrimination resolution compared to the other examined regions and presented almost the same topology to that obtained from the concatenated matrix. Thus, the *cox*3–*nad*3 region may be a good alternative to the concatenated approach for typing, better representing the general mt variation. Thirdly, the *cob* intronic region showed the lowest discriminatory information but adds value by providing evidence of mt intron variability, a feature also described as a contributor to fungal mt genome length variation (De Chiara *et al.* 2020; Megarioti and Kouvelis 2020). It is interesting that the second ID intron, at position 429 of the *cob* gene, can also be found in fungal species belonging to Pezizomycotina (Cinget and Bélanger 2020), indicating that this intron seems to be an ancestral element due to its existence in two divergent fungal phyla (Basidiomycota and Ascomycota). Moreover, this intron was found to be highly mobile. Cinget and Bélanger (2020) showed that this intron was related to an adaptive fungicide resistance as a result of a mutation in the *cob* gene and it has a regulatory role in *cob* gene maturation (Grasso *et al.* 2006; Vallieres *et al.* 2011).

The mt-based trees provide different discriminatory results when compared with the IGS-based topology and several interesting conclusions may be extracted for strain genotyping. Based on the IGS phylogeny, strains CBS7982, CBS9369 and CBS9589 of group H are phylogenetically distant from strains of group A1/A2 (Fig. 2). On the contrary, a mt-based phylogeny showed that these strains are more closely related to strains of group A1/A2, rather than the other members of IGS group H (mt concatenated groups III2 and III3). This observation may be indicative of mt recombination in these groups. For the 15 protein coding gene phylogenetic approaches, CBS 7982 (III2/H) also clustered with III3/A2 strain CBS 14141, supporting this finding (Fig. 5). A similar phenomenon is observed for IGS G strain CBS5332, which belongs to the III3 mt concatenated type, a group otherwise almost exclusively containing strains belonging to nuclear IGS type A1/A2. For this strain, no mt genome sequence was available.

The strains analyzed in this work may be divided in three main categories according to their host and pathogenicity: The first group is pathogenic to animal skin, the second is pathogenic to human skin, and the third consisted of deep-seated strains (Table 1). IGS-based phylogeny clusters most animal pathogenic strains (group B) as members of a sister clade to most deep-seated strains (group A; Fig. 2). On the contrary, phylogeny-based on the mt concatenated data clearly separates the strains infecting animals (group I) from the human deep-seated strains (groups III1 and III3; Fig. 4). This discrimination based on the mt concatenated matrix may be attributed to the differential evolution of the intergenic mt regions which are usually under neutral selection (Kimura 1991; Bartelli *et al.* 2013), while the IGS region contains several regulatory elements (Pantou, Mavridou and Typas 2003) which might diminish changes throughout evolution. Thus, the IGS region shows less discrimination and sisterhood of the animal skin and deep-seated pathogenic strains. Based on a limited number of samples, this seems also the case when considering the 15 mt protein coding genes (Fig. 5). When comparing mt phylogeny of the concatenated mt target regions (Fig. 4) with the phylogeny based on all protein coding genes (Fig. 5), some differences can be observed. Specifically, mt groups II3 and III3 seem to be scattered over the tree based on the protein coding genes.

One of the most striking observations is that previous suggestions of a specific genotype in deep-seated infections (Theelen *et al.* 2001; Gupta *et al.* 2004) is also observed in this study using IGS and mt target sequence data. Ongoing research evaluating a large set of over 150 clinical isolates from various patients and geographies further confirms this finding (B. Theelen, unpublished observations). This genotype is not exclusive for deep-seated isolates, but deep-seated isolates are not observed in any other genotype. There is one exception, pertaining to strains MAL26 and MAL32, which are both neonatal strains originating from the same patient, belonging to genotype IGS G, whereas all other neonatal and deep-seated isolated belong to genotype IGS A1/A2. Both isolates belong to mt group III1. Interestingly, the limited genotype variation for deep seated isolates found for IGS and the mt target regions does not seem to be supported when considering the mt protein coding gene phylogeny (Fig. 5). This may be attributed to the conservation of gene sequences, since they are crucial to cell surviving and thus, difficult to introduce any changes.

Overall, the use of various mt regions for typing of *M. furfur* can be debated, depending on purpose and functional clustering. Future analysis of a larger set of strains, including from clinical origin, will determine which region(s) will provide the best resolution for intraspecies discrimination. Presently, it can be concluded that the use of the *cox3–nad3* mt intergenic region presents the most promising typing prospect for isolates of *M. furfur* because it best represents the concatenated mt variation, and the general clustering is similar to that based on IGS, but it provides additional discrimination within clusters. Both IGS and *cox3–nad3* domains may provide further insights into the different evolutionary processes of nuclear and mt DNA. Based on the limited number of samples and their host backgrounds, at this stage it is difficult to draw conclusions regarding genotype to host specificity or pathogenicity, but our comparative mitogenome analysis indicated several very promising regions for strain typing. Our study adds to the general knowledge of mt genome organization in the genus *Malassezia*, and more specifically in *M. furfur*, a relevant species involved in skin disease and emerging in BSIs. Various heterogenous mt regions were identified, providing promising multicopy targets for future typing studies, diagnostic assay development and evolutionary studies.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSYR online.

ACKNOWLEDGMENTS

The authors wish to thank Claudia Cafarchia of the Università degli Studi 'Aldo Moro', Bari, Italy for providing MAL18, MAL24, MAL26, MAL32 and CD864 strains of *Malassezia furfur* and Guangxi Wu, Zymo Research, USA for his support in initial read assembly. 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. Thomas Dawson is supported by the ASTAR Industry Alignment Fund (H18/01/a0/016). VNK would like to thank HEAL-Link, for financially supporting the publication of this article in OA mode.

Conflict of interest. None declared.

REFERENCES

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Amend A. From dandruff to deep-sea vents: Malassezia-like fungi are ecologically hyper-diverse. *PLoS Pathogens* 2014;**10**:e1004277.

Archibald JM. Endosymbiosis and eukaryotic cell evolution. *Curr Biol* 2015;**25**:R911–21.

Aykut B, Pushalkar S, Chen R *et al.* The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* 2019;**574**:264–7.

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.

Bartelli TF, Ferreira RC, Colombo AL *et al.* Comparative genomics of *Candida albicans* mitochondria reveals non-coding regions under neutral evolution. *Infect Genet Evol* 2013;**14**:302–12.

Batra R, Boekhout T, Guého E *et al. Malassezia baillon*, emerging clinical yeasts. *FEMS Yeast Res* 2005;**5**:1101–13.

Belfort M, Derbyshire V, Parker MM *et al.* Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M *et al.* (eds). *Mobile DNA II*. Washington, DC: ASM Press, 2002, 761–83.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20.

Brooks R, Brown L. Systemic infection with *Malassezia furfur* in an adult receiving long-term hyperalimentation therapy. *J Infect Dis* 1987;**156**:410–1.

Burland TG. DNASTAR's lasergene sequence analysis software. *Methods Mol Biol* 2000;**132**:71–91.

Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* 2019;**1962**:1–14.

Chen IT, Chen CC, Huang HC *et al. Malassezia furfur* emergence and Candidemia trends in a neonatal intensive care unit during 10 years: the experience of fluconazole prophylaxis in a single hospital. *Adv Neonatal Care* 2020;**20**:E3–8.

Cinget B, Bélanger RR. 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* 2020;**17**:1252–60.

Darriba D, Taboada GL, Doallo R *et al.* jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012;**9**:772.

De Chiara M, Friedrich A, Barré B *et al.* Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biol* 2020;**18**:49.

Deng Y, Hsiang T, Li S *et al.* 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* 2018;**9**:2079.

Findley K, Oh J, Yang J *et al.* Topographic diversity of fungal and bacterial communities in human skin. *Nature* 2020;**498**:367–70.

Formey D, Molès M, Haouy A *et al.* Comparative analysis of mitochondrial genomes of *Rhizophagus irregularis* - syn. *Glomus irregulare* - reveals a polymorphism induced by variability generating elements. *New Phytol* 2012;**196**:1217–27.

Foury F, Roganti T, Lecrenier N *et al.* The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett* 1998;**440**:325–31.

Freel KC, Friedrich A, Schacherer J. Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res* 2015;**15**:fov023. DOI: 10.1093/femsyr/fov023.

Gaitanis G, Bassukas ID, Velegraki A. The range of molecular methods for typing Malassezia. *Curr Opin Infect Dis* 2009;**22**:119–25.

Gaitanis G, Robert V, Velegraki A. Verifiable single nucleotide polymorphisms of the internal transcribed spacer 2 region for the identification of 11 Malassezia species. *J Dermatol Sci* 2006;**43**:214–7.

Ghikas DV, Kouvelis VN, Typas MA. 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* 2010;**10**:174.

Ghikas DV, Kouvelis VN, Typas MA. 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* 2006;**185**:302–8933.

Gioti A, Nystedt B, Li W *et al*. Genomic insights into the atopic eczema-associated skin commensal yeast *Malassezia sympodialis*. *mBio* 2013;**4**:e00572–12.

Grasso V, Palermo S, Sierotzki H *et al*. Cytochrome b gene structure and consequences for resistance to Qo inhibitor fungicides in plant pathogens. *Pest Manag Sci* 2006;**62**:465–72.

Guého E, Boekhout T, Ashbee HR *et al*. The role of Malassezia species in the ecology of human skin and as pathogens. *Med Mycol* 1998;**36**:220–9.

Guého-Kellermann E, Boekhout T, Begerow D. Biodiversity, phylogeny and ultrastructure. In: Boekhout T, Guého E, Mayser P *et al*. (eds). *Malassezia and the Skin*. Berlon Heidelberg: Springer Verlag, 2010, 17–63.

Guillot J, Bond R. Malassezia yeasts in veterinary dermatology: an updated overview. *Front Cell Infect Microbiol* 2020;**10**:79.

Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;**52**:696–704.

Gupta AK, Boekhout T, Theelen B *et al*. Identification and typing of Malassezia species by amplified fragment length polymorphism and sequence analyses of the internal transcribed spacer and large-subunit regions of ribosomal DNA. *J Clin Microbiol* 2004;**42**:4253–60.

Honnavar P, Dogra S, Handa S *et al*. Molecular identification and quantification of Malassezia species isolated from pityriasis versicolor. *Ind Dermatol Online J* 2020;**11**:167–70.

Iatta R, Cafarchia C, Cuna T *et al*. Bloodstream infections by Malassezia and Candida species in critical care patients. *Med Mycol* 2014;**52**:264–9.

Kaneko T, Murotani M, Ohkusu K *et al*. Genetic and biological features of catheter-associated *Malassezia furfur* from hospitalized adults. *Med Mycol* 2012;**50**:74–80.

Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019;**20**:1160–6.

Kimura M. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet* 1991;**66**:367–86.

Korovesi AG, Ntertilis M, Kouvelis VN. Mt-rps3 is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. *Mol Phylogenet Evol* 2018;**127**:74–86.

Kortsinoglou AM, Korovesi AG, Theelen B *et al*. The mitochondrial intergenic regions nad1-cob and cob-rps3 as molecular identification tools for pathogenic members of the genus Cryptococcus. *FEMS Yeast Res* 2019;**19**:foz077.

Kortsinoglou AM, Saud Z, Eastwood DC *et al*. The mitochondrial genome contribution to the phylogeny and identification of Metarhizium species and strains. *Fungal Biol* 2020;**124**:845–53.

Kouvelis VN, Ghikas DV, Typas MA. 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. *Fung Genet Biol* 2004;**41**:930–40.

Kouvelis VN, Sialakouma A, Typas MA. Mitochondrial gene sequences alone or combined with ITS region sequences provide firm molecular criteria for the classification of Lecanicillium species. *Mycol Res* 2008;**112**:829–44.

Kuraku S, Zmasek CM, Nishimura O *et al*. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res* 2013;**41**:W22–8.

Limon JJ, Tang J, Li D *et al*. Malassezia is associated with Crohn's disease and exacerbates colitis in mouse models. *Cell Host Microbe* 2019;**25**:377–88.

Liu X, Wu X, Tan H *et al*. Large inverted repeats identified by intra-specific comparison of mitochondrial genomes provide insights into the evolution of *Agrocybe aegerita*. *Comput Struct Biotechnol J* 2020;**18**:2424–37.

Lorch JM, Palmer JM, Vanderwolf KJ *et al*. *Malassezia vespertilionis* sp. nov.: a new cold-tolerant species of yeast isolated from bats. *Persoonia* 2018;**41**:56–70.

Megarioti AM, Kouvelis VN. The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG), *Genome Biol Evol* 2020;**12**:1337–54.

Nieuwenhuis M, van de Peppel LJJ, Bakker FT *et al*. Enrichment of G4DNA and a large inverted repeat coincide in the mitochondrial genomes of termitomyces. *Genome Biol Evol* 2019;**11**:1857–69.

O'Donnell K, Cigelnik E, Weber NS *et al*. Phylogenetic relationships among ascomycetous truffles and the true and false morels inferred from 18S and 28S rDNA sequence analysis. *Mycologia* 1997;**89**:48–65.

Pantou MP, Mavridou A, Typas MA. IGS sequence variation, group-I introns and the complete nuclear ribosomal DNA of the entomopathogenic fungus *Metarhizium*: excellent tools for isolate detection and phylogenetic analysis. *Fungal Genet Biol* 2003;**38**:159–74.

Rhimi W, Theelen B, Boekhout T *et al*. *Malassezia* spp. yeasts of emerging concern in fungemia. *Front Cell Infect Microbiol* 2020;**10**:370.

Ronquist F, Huelsenbeck JP. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;**19**:1572–4.

Saunte DML, Gaitanis G, Hay RJ. *Malassezia*-Associated skin diseases, the use of diagnostics and treatment. *Front Cell Infect Microbiol* 2020;**10**:112.

Spatz M, Richard ML. Overview of the potential role of *Malassezia* in gut health and disease. *Front Cell Infect Microbiol* 2020;**10**:201.

Sugita T, Kodama M, Saito M *et al*. Sequence diversity of the intergenic spacer region of the rRNA gene of *Malassezia globosa* colonizing the skin of patients with atopic dermatitis and healthy individuals. *J Clin Microbiol* 2003;**41**:3022–7.

Sugita T, Nakajima M, Ikeda R *et al*. Sequence analysis of the ribosomal DNA intergenic spacer 1 regions of Trichosporon species. *J Clin Microbiol* 2002;**40**:1826–30.

Sugita T, Takeo K, Hama K *et al*. DNA sequence diversity of intergenic spacer I region in the non-lipid-dependent species *Malassezia pachydermatis* isolated from animals. *Med Mycol* 2005;**43**:21–6.

Theelen B, Cafarchia C, Gaitanis G *et al*. Malassezia ecology, pathophysiology, and treatment [published correction appears in Med Mycol. 2019 Apr 1;57(3):e2]. *Med Mycol* 2018;**56**:S10–25.

Theelen B, Silvestri M, Guého E *et al*. Identification and typing of Malassezia yeasts using amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and denaturing gradient gel electrophoresis (DGGE). *FEMS Yeast Res* 2001;**1**:79–86.

Untergasser A, Cutcutache I, Koressaar T *et al*. Primer3: new capabilities and interfaces. *Nucleic Acids Res* 2012;**40**:e115.

Valach M, Farkas Z, Fricova D *et al*. Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Res* 2011;**39**:4202–19.

Vallières C, Trouillard M, Dujardin G *et al*. Deleterious effect of the Qo inhibitor compound resistance-conferring mutation G143A in the intron-containing cytochrome b gene

and mechanisms for bypassing it. *Appl Environ Microbiol* 2011;**77**:2088–93.

Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics* 2003. DOI: 10.1002/0471250953.bi0604s00. Chapter 6:Unit 6.4.

Wolters JF, Chiu K, Fiumera HL. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genomics* 2015;**16**:451.

Wu G, Zhao H, Li C *et al*. Genus-wide comparative genomics of Malassezia delineates its phylogeny, physiology, and niche adaptation on human skin. *PLoS Genet* 2015;**11**:e1005614.

Zhang H, Zhang R, Ran Y *et al*. Genetic polymorphism of *Malassezia furfur* isolates from Han and Tibetan ethnic groups in China using DNA fingerprinting. *Med Mycol* 2010;**48**: 1034–8.

Zhang S, Hao AJ, Zhao YX *et al*. Comparative mitochondrial genomics toward exploring molecular markers in the medicinal fungus *Cordyceps militaris*. *Sci Rep* 2017;**7**:40219.

Zhu Y, Engström PG, Tellgren-Roth C *et al*. Proteogenomics produces comprehensive and highly accurate protein-coding gene annotation in a complete genome assembly of *Malassezia sympodialis*. *Nucleic Acids Res* 2017;**45**:2629–43.