HAL open science

# Comparative genomics applied to Mucor species with different lifestyles

Annie Lebreton, Erwan Corre, Jean-Luc Jany, Loraine Brillet-Guéguen, Carlos Pèrez-Arques, Victoriano Garre, Misharl Monsoor, Robert Debuchy, Christophe Le Meur, Emmanuel Coton, et al.

## To cite this version:

Annie Lebreton, Erwan Corre, Jean-Luc Jany, Loraine Brillet-Guéguen, Carlos Pèrez-Arques, et al.. Comparative genomics applied to Mucor species with different lifestyles. BMC Genomics, 2020, 21 (1), pp.135. 10.1186/s12864-019-6256-2 . hal-02481326

**HAL Id: hal-02481326**

**https://hal.science/hal-02481326v1**

Submitted on 10 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Lebreton et al. BMC Genomics (2020) 21:135
https://doi.org/10.1186/s12864-019-6256-2

BMC Genomics

RESEARCH ARTICLE Open Access

# Comparative genomics applied to *Mucor* species with different lifestyles

Annie Lebreton[1], Erwan Corre[2], Jean-Luc Jany[1], Loraine Brillet-Guéguen[2,3], Carlos Pèrez-Arques[4], Victoriano Garre[4], Misharl Monsoor[2], Robert Debuchy[5], Christophe Le Meur[1], Emmanuel Coton[1], Georges Barbier[1] and Laurence Meslet-Cladière[1*]

## Abstract

**Background:** Despite a growing number of investigations on early diverging fungi, the corresponding lineages have not been as extensively characterized as *Ascomycota* or *Basidiomycota* ones. The *Mucor* genus, pertaining to one of these lineages is not an exception. To this date, a restricted number of *Mucor* annotated genomes is publicly available and mainly correspond to the reference species, *Mucor circinelloides,* and to medically relevant species. However, the *Mucor* genus is composed of a large number of ubiquitous species as well as few species that have been reported to specifically occur in certain habitats. The present study aimed to expand the range of *Mucor* genomes available and identify potential genomic imprints of adaptation to different environments and lifestyles in the *Mucor* genus.

**Results:** In this study, we report four newly sequenced genomes of *Mucor* isolates collected from non-clinical environments pertaining to species with contrasted lifestyles, namely *Mucor fuscus* and *Mucor lanceolatus*, two species used in cheese production (during ripening), *Mucor racemosus*, a recurrent cheese spoiler sometimes described as an opportunistic animal and human pathogen, and *Mucor endophyticus*, a plant endophyte. Comparison of these new genomes with those previously available for six *Mucor* and two *Rhizopus* (formerly identified as *M. racemosus*) isolates allowed global structural and functional description such as their TE content, core and species-specific genes and specialized genes. We proposed gene candidates involved in iron metabolism; some of these genes being known to be involved in pathogenicity; and described patterns such as a reduced number of CAZymes in the species used for cheese ripening as well as in the endophytic isolate that might be related to adaptation to different environments and lifestyles within the *Mucor* genus.

**Conclusions:** This study extended the descriptive data set for *Mucor* genomes, pointed out the complexity of obtaining a robust phylogeny even with multiple genes families and allowed identifying contrasting potentially lifestyle-associated gene repertoires. The obtained data will allow investigating further the link between genetic and its biological data, especially in terms of adaptation to a given habitat.

**Keywords:** Early divergent fungi, Genome, Adaptation, CAZymes, Peptidases, Iron uptake

## Background

The *Mucor* genus belongs to the most prominent order of the *Mucorales*, a phylogenetically ancient group of fungi pertaining to the "early diverging fungi" [1]. From the first microscopic observation of a *Mucor* specimen in 1665 up until now, several hundreds of potential *Mucor* species have been reported [2]. *Mucor* species are common and predominantly saprotrophs [3]. These ubiquitous microorganisms may colonize multiple and contrasted environments from dungs or dead plant materials to plant and animal tissues. Members of the *Mucor* genus have an ambivalent impact on human activities. Regarding their negative impact, some *Mucor* species, in particular the thermotolerant species *Mucor indicus, Mucor ramosissimus* and seven members of the *Mucor circinelloides* complex [4] have been shown to be

* Correspondence: laurence.meslet@univ-brest.fr
[1]Univ Brest, Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, F-29280 Plouzané, France
Full list of author information is available at the end of the article

© The Author(s). 2020 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

human and animal pathogens responsible for mucormycosis [2]. Mucormycosis has been recently described as the third most common angioinvasive fungal infection and can lead to death [5]. Another negative impact concerns the ability of different species of the genus to spoil raw and transformed foods and feeds [6]. On the contrary, some *Mucor* species have an important biotechnological potential thanks to their high growth rates in a large range of temperatures [7], existence of a yeast state in certain *Mucor* spp. [8], and high proteolytic and lipolytic enzymatic activities [9], making them good candidates for biotechnology. Interestingly, some species are also used in food manufacturing of Asian fermented food production (such as ragi, tempeh, furu or mureha) or for French cheese ripening (such as Tomme or Saint-Nectaire) [2].

The increasing number of infections associated with *Mucor* species, as well as the biotechnological potential of the genus, have led to a large effort to better know these fungi. In this context, Vongsangnak et al. proposed a metabolic network of the oleaginous strain *Mucor circicinelloides* WJ11 [10, 11], Corrochano et al. shed new light on *Mucor* sensory perception [12], and multiple genes potentially involved in virulence were investigated and discovered in *Mucor* spp. [13–21]. Following the annotation of the first *Mucor* genome sequence (*Mucor circinelloides* CBS 277.49), researches on *Mucor* benefited and will continue to benefit from different sequencing projects including the Zygolife initiative (http://zygolife.org/home/) which aims to provide a better phylogenetical classification to the formerly called Zygomycetes which include the *Mucor* genus (see [1]).

This phylogenetical classification appears to be challenging as stated by inconsistencies among previous works; e.g. *Mucor indicus* CDC-B7402 placement was modified between the phylogeny of Álvarez et al. and that of Whalter et al., *Mucor endophyticus* CBS 385–95 placement was modified between Whalter et al. and Lebreton et al. respective studies [22–24]. Moreover, the uncertain taxonomic assignment of some *Mucor* isolates used in published studies may lead to confusion. For example, following genomic studies, the 97–1192 isolate was reassigned from *Mucor racemosus* to *Rhizopus oryzae* and isolate CDC-B9738 (initially *Rhizopus microsporus)* was consecutively assigned to *Mucor racemosus* by Chibucos et al. and more recently reassigned to *R. microsporus* by Gryganskyi et al. [13, 25]. As stated by Gryganskyi et al., the closely related genus *Rhizopus* cannot be deciphered with a single or even a handful of gene families [25]. However, the range of *Mucor* genome sequences exploited is limited and those available with annotations even scarcer. Furthermore, these genome sequencing projects are mainly limited to *Mucor* species with a biotechnological or pathogenic potential. Indeed, at the beginning of this study, only six *Mucor* annotated genomes were freely available, five of them corresponding to isolates pertaining to the *M. circinelloides* complex.

The present study aimed to use comparative genomics to identify potential genomic imprints of adaptation to different environments and lifestyles in the *Mucor* genus. To do so, four genomes, corresponding to *Mucor fuscus* UBOCC-A-109160 and *Mucor lanceolatus* UBOCC-A-109153 (used in cheese production, during ripening), *Mucor racemosus* UBOCC-A-109155 (a cheese spoiler sometimes described as an opportunistic animal and human pathogen) [26] and *Mucor endophyticus* CBS 385–95 (a plant endophyte [27]), were sequenced, annotated and compared to those of six publicly available *Mucor* and two *Rhizopus* (formerly identified as *Mucor*) isolates.

## Results

### Genome description

#### *Genome sequences and assembly*

The genomes of *M. fuscus* UBOCC-A-109160, *M. lanceolatus* UBOCC-A-109153, *M. racemosus* UBOCC-A-109155 and *M. endophyticus* CBS 385–95 were sequenced, assembled and annotated in the context of this study. Their genome features were compared to eight previously sequenced and annotated genomes from *Mucor* isolates ($n = 6$) and *Rhizopus* isolates formerly identified as *Mucor* spp. ($n = 2$) [12, 13, 28] (Table 1). Due to the recent reassignment of the *Mucor circinelloides* complex members to distinct species [4], isolates identified as *M. circinelloides* were renamed according to their current phylogenetic placement: *Mucor circinelloides* CBS 277.49 was renamed into *Mucor lusitanicus* CBS 277.49, *Mucor circinelloides* CDC-B5328 into *Mucor velutinosus* CDC-B5328 and *Mucor circinelloides* NBRC 6742 (synonym of *Mucor ambiguus* NBRC 6742) into *Mucor griseocyanus* NBRC 6742. For simplicity, with the exception of *M. circinelloides* and *R. microsporus* representatives (two isolates were analysed both for *M. circinelloides* and *R. microsporus* species), the different isolates investigated will, hereafter, be named by their species name.

The number of scaffolds in *M. lusitanicus* assembly was the lowest with 21 scaffolds. *M. endophyticus* was the second less fragmented assembly with 159 scaffolds. The 10 other assemblies were composed of 470 to 3888 scaffolds (Table 1).

Despite these differences in genome fragmentation, at least 95% of the 290 single copy fungal orthologous genes searched by BUSCO were found complete in all genomes. It is worth noting that respectively 52 and 87% of the searched genes were found duplicated in *Rhizopus microsporus* CDC-B9645 and *Rhizopus microsporus* CDC-B9738 genomes whereas, in the other genomes, duplicated genes only represented 17 to 26% of the searched genes. The two *Rhizopus microsporus* isolates also exhibited the highest genome size with 65 Mb and 75 Mb for CDC-B9645 and CDC-B9738 isolates, respectively, while the average size for

**Table 1** Genome assembly features and structural annotation of the studied *Mucor* and *Rhizopus* isolates. The four isolates sequenced in the context of this work are in red. IG: intergenic regions, TE: transposable elements

| | *Mucor endophyticus* CBS 385–95 | *Mucor fuscus* UBOCC-A-109160 | *Mucor lanceolatus* UBOCC-A-109153 | *Mucor racemosus* UBOCC-A-109155 | *Mucor griseocyanus* NBRC 6742 | *Mucor circinelloides* 1006PhL | *Mucor circinelloides* CDC-B8987 | *Mucor lusitanicus* CBS 277.49 | *Mucor velutinosus* CDC-B5328 | *Mucor indicus* CDC-B7402 | *Rhizopus microsporus* CDC-B9645 [a] | *Rhizopus microsporus* CDC-B9738 [a] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Genome size (Mb) | 35,00 | 40,60 | 43,42 | 46,92 | 40,74 | 36,35 | 36,77 | 36,57 | 35,61 | 39,62 | 64,72 | 74,89 |
| # Scaffolds > 1000 bp | 159 | 3819 | 1531 | 3506 | 1283 | 470 | 1022 | 21 | 1016 | 798 | 3888 | 2676 |
| %GC | 34 | 36 | 34 | 32 | 32 | 37 | 39 | 42 | 40 | 36 | 32 | 33 |
| %N | 2 | 0 | 5 | 0 | 22 | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| Maximum scaffold (kb) | 4527 | 142 | 681 | 162 | 539 | 664 | 340 | 6050 | 424 | 831 | 518 | 3513 |
| N50 (kb) | 1957 | 27 | 142 | 25 | 114 | 141 | 92 | 4318 | 79 | 271 | 38 | 90 |
| N90 (kb) | 575 | 4 | 13 | 7 | 26 | 51 | 23 | 1075 | 21 | 33 | 10 | 14 |
| # Genes | 11,799 | 12,571 | 10,924 | 11,604 | 11,726 | 12,410 | 10,437 | 11,936 | 9997 | 11,703 | 15,153 | 21,229 |
| Gene density (gene/Mb) | 337 | 310 | 252 | 247 | 288 | 341 | 284 | 326 | 281 | 295 | 234 | 283 |
| Ave.gene length (bp) | 1677 | 1542 | 1677 | 1633 | 1616 | 1491 | 1609 | 1408 | 1608 | 1491 | 1517 | 1543 |
| Ave. exon frequency | 4,54 | 4,33 | 4,55 | 4,27 | 4,05 | 3,83 | 4,09 | 3,73 | 4,11 | 3,99 | 4,00 | 4,16 |
| Ave. exon length (bp) | 313 | 302 | 301 | 335 | 340 | 340 | 336 | 311 | 335 | 316 | 324 | 318 |
| % genes with introns | 86 | 82 | 87 | 84 | 81 | 82 | 83 | 81 | 84 | 83 | 81 | 83 |
| Ave. intron length (bp) | 82 | 76 | 95 | 72 | 77 | 66 | 76 | 91 | 74 | 76 | 74 | 69 |
| Ave. IG size (bp) | 1293 | 1284 | 2040 | 1791 | 1673 | 1374 | 1677 | 1691 | 1758 | 1760 | 2151 | 1713 |
| # Prot. predicted | 11,437 | 12,310 | 10,636 | 11,269 | 11,343 | 12,227 | 10,190 | 11,709 | 9716 | 11,390 | 14,656 | 20,723 |
| %TE coverage | 5,08 | 14,28 | 22,8 | 36,8 | 6,03 | 15,2 | 22,18 | 23,44 | 24,43 | 16,51 | 38,47 | 34,46 |
| Genome references | | | | | | [29] | [13] | [12] | [13] | [13] | [13] | |

[a]Initially assigned to *Mucor racemosus* [13] but later reassigned to *R. microsporus* [25]

the other geomes is approximately 39 Mb. The endophytic species *M. endophyticus*, had the smallest genome size (35 Mb) (Table 1).

### *Genome annotation*

The number of predicted genes was in accordance with genome sizes. In *Mucor* species, the number of predicted genes fluctuated from 9997 to 12,571, i.e. a gene density ranging from 247 to 341 genes/Mb, while 15,153 and 21, 229 genes were predicted for *R. microsporus* CDC-B9645 and CDC-B9738, i.e. a gene density of 234 and 283 genes/Mb, respectively (Table 1).

Gene characteristics were well conserved among the *Mucor* and *Rhizopus* genomes: the average gene length was 1568 bp, 83% of genes had predicted introns, genes had an average frequency of 4.1 exons and average intron size was approximately 60 bp.

Noteworthy, within all genomes, the intergenic distance was variable: 25% of intergenic regions were shorter than ~ 300 bp whereas the largest intergenic regions exceeded 20 kb. When repeated elements were taken into account for the analysis, regions up to 15 kb with neither gene nor repeated elements were still detected (Fig. 1).

## Genome comparisons

### *Phylogenomic reconstruction*

The 181,601 proteins predicted from the 10 *Mucor*, three *Rhizopus* and the *Phycomyces blakesleeanus* genomes were grouped in 20,588 orthogroups. Among them, 4240 orthogroups were composed of predicted proteins belonging to all genomes while 64 were composed of single copy orthologs. Among the 64-single copy orthogroups, 51 were selected to reconstruct the species tree using Maximum Likelihood and Bayesian methods, *R. delemar* and *P. blakesleeanus* being used as outgroups (Fig. 2a). The present study confirmed the recently described phylogenetic positions of the former *M. circinelloides* formae: *M. griseocyanus*, *M. lusitanicus* and *M. velutinosus* as distinct clades from *M. circinelloides* [4]. The placement of *M. indicus* was altered compared to the topology obtained by Whalter et al. but concurred with the topology published by Álvarez et al. [22, 24]. Similarly, the placement of *R. microsporus* CDC-B9645 and *R. microsporus* CDC-B9738 was concordant with the topology of Chibucos et al. which identified these two isolates as *M. racemosus*, but clearly differentiated from the one by Gryganskyi et al. in which the two isolates were clustered with *R. microsporus* isolates (and renamed accordingly) [13, 25]. This phylogenetic position was not affected when four previously sequenced *R. microsporus* were added to the phylogenomic reconstruction (Additional file 1: Figure S1). These results clearly raise questions concerning their actual position and the genetic bases associated with these incongruences. The other studied isolates had concordant phylogenetic placements with previously published studies [13, 22, 23, 25].

From the biological point of view, *M. fuscus* and *M. lanceolatus* that are close in the tree, have mainly been

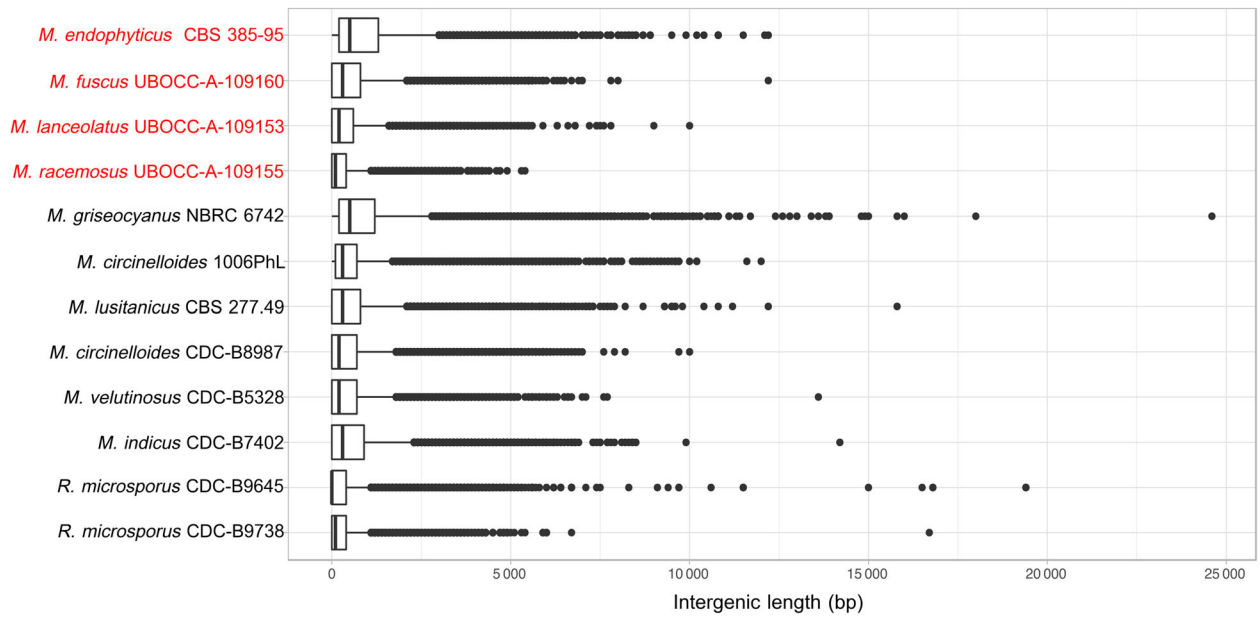


**Fig. 1** Repartition of the intergenic length within the 12 studied *Mucor* and *Rhizopus* genomes. The four isolates sequenced in the context of this study are in red. The boxes represent intergenic length harbored by 25 to 75% (sorted by length) of the intergenic regions, the line within the box represent the median length, dots represent each value corresponding to the 10% longest intergenic units represented

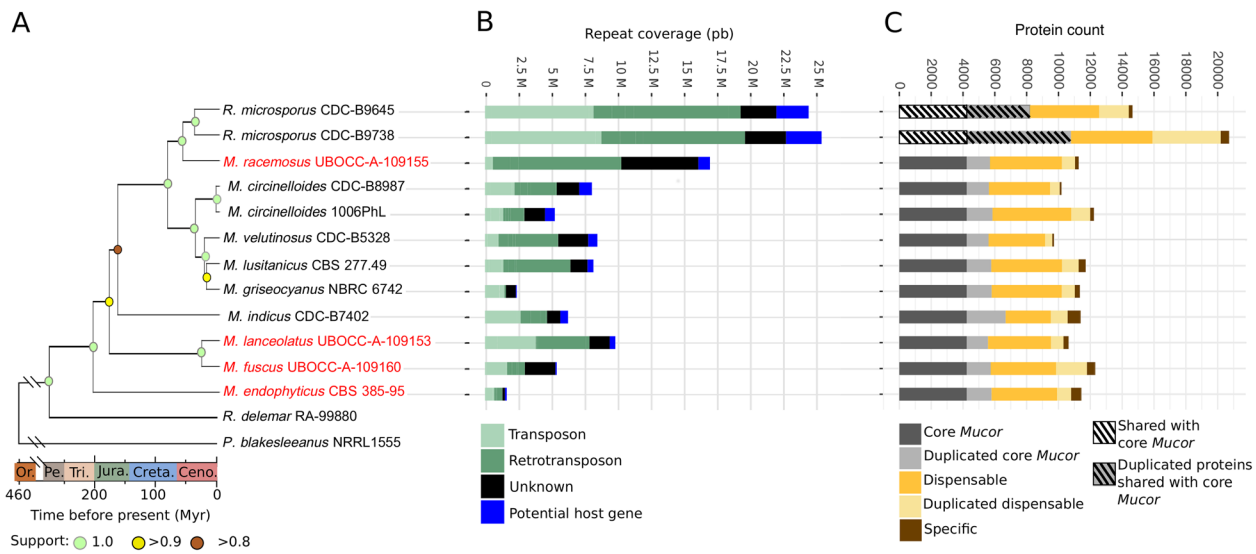


**Fig. 2** Representation of the phylogenomic tree, repetitions coverage and gene number for the genomes pertaining to the studied *Mucor* and *Rhizopus* isolates (initially assigned to *M. racemosus*). **a** Phylogenomic tree of the studied isolates reconstructed with RAxML based on 51 single copy ortholog families using *R. delemar* RA-99880 and *P. blakesleeanus* NRRL1555 as outgroup and calibrated on time scale with r8S program. The branch length corresponds to the estimated time before present. Color dots represent bootstrap support. The four isolates sequenced in this study are in red. **b** Representation of the repetitions coverage in each genome. Unknown: potential transposable element that could not be identified. **c** Representation of the protein numbers in the studied isolates. Duplicated proteins count represent extra copy number of each gene family

encountered in cheeses, the basal singleton *M. endophyticus* has only been described as a wheat leaf endophyte, whereas the other *Mucor* species, grouped in the same clade, are usually described as potential human pathogenic species.

### *Transposable elements*

The studied genomes contained contrasting transposable element (TE) coverages (Fig. 2b). The plant endophyte, *M. endophyticus*, held the lowest TE coverage (5%) whereas the ubiquitous *M. racemosus* contained the highest TE coverage (37%). Even between close species, TE coverage and composition differed notably, e.g. between *M. lanceolatus* and *M. fuscus* or between *M. griseocyanus* and the other species from the *M. circinelloides* complex (*M. circinelloides*, *M. lusitanicus* and *M. velutinosus*). *M. lanceolatus* showed higher TE coverage than *M. fuscus* (23 and 14%, respectively) for both transposons and retrotransposons but the predicted elements in *M. lanceolatus* were mainly non-autonomous (i.e. not including all the domains necessary for their transposition). On the contrary, almost all *M. fuscus* predicted TE were autonomous, the main represented category being terminal inverted repeat (TIR) but few retrotransposons, long interspersed nuclear elements (LINE) and long terminal repeat (LTR), were also detected (Additional file 2: Figure S2).

Among members of the *M. circinelloides* complex, three species displayed a similar pattern in terms of TE composition and, except for one of the *M. circinelloides* isolates (1006PhL), in terms of coverage. The fourth species, *M. griseocyanus,* displayed striking differences from the other species. Indeed, the TE coverage of *M. griseocyanus* was reduced by at least twofold compared to the other members (Fig. 2b). Furthermore, the main TE category was identified as belonging to the helitron order, a marginal category in the three other members of the clade.

### *Predicted proteins repartition*

The predicted protein repertoires encoded by the 14 *Mucoromycota* investigated were compared with one another. This led to the identification of a set of core predicted proteins shared by all studied *Mucor* species and to the determination of the proteins also occurring in the studied *R. microsporus* isolates. Dispensable proteins (shared by at least two species) and species-specific proteins (only found in a single species) were also identified (Fig. 2c). For each category, duplicated proteins were identified. Despite the phylogenetic divergence between species, at least one third of the predicted proteins for each species were part of the core predicted proteome and the number of species-specific proteins was restricted.

As suggested by BUSCO results, most of *R. microsporus* predicted proteins were duplicated (Fig. 2c). This result explained the relatively low number of single copy core proteins gathered for phylogeny reconstruction despite a core *Mucor* proteome of approximately 6000 proteins.

### *Evolution of gene families across phylogeny*

When using the whole genome dataset, CAFE-, DupliPhyML- and Notung-based analyses yielded non-concordant results (data not shown) with inconsistent placements of expansion/contraction events within the cluster encompassing *R. microsporus* CDC-B9738, *R. microsporus* CDC-B9645 and *M. racemosus*. This behaviour was interpreted as a side effect of a putative whole genome duplication or hybridization compatible with the observations done on the two *R. microsporus* genomes. The phylogenetic placement of these two isolates within the genus *Mucor* being also questioned, we decided to remove them from the CAFE analysis. CAFE identified 44 rapidly evolving gene families on the *M. lanceolatus*/*M. fuscus* branch (pertaining to the two species associated to cheese ripening). Among these families, two were associated to secondary metabolism, namely an acyl-CoA synthetase and a cytochrome p450 encoding gene families, both with reduced number of genes in cheese ripening species (*M. fuscus* and *M. lanceolatus*) compared to other species. A cysteine hydrolase gene family was also reduced in the two cheese-associated species. Another family (less conserved) with genes identified as encoding putative transcriptional activators of glycolytic enzyme was expanded in the latter species genomes. Other families were either unknown or similar to TE sequences.

In the *M. endophyticus* endophyte, at the node separating cheese-associated species from pathogenic species and at the node separating *M. indicus* from other species, 49, 4 and 9 gene families were considered as rapidly evolving, respectively. However, these gene families were either of unknown function or similar to TE sequences.

### *Genes coding for CAZymes, peptidases and small secreted proteins*

As different CAZyme and peptidase repertoires are associated to different lifestyles, comparison of the annotations of CAZyme and peptidase genes annotation of *M. lusitanicus* were publicly available in the MycoCosm database (https://mycocosm.jgi.doe.gov/mycocosm/home). However, for the sake of the comparison, those annotations were performed again in this study following the same pipeline as the other studied genomes. This allowed to show that, for *M. lusitanicus*, the number of peptidase genes was found higher (351 against 304) and the number of CAZymes lower (229 against 339) in this study compared to annotation performed by the Join Genome Institute and displayed in the MycoCosm database. Taking this difference into account, the total numbers of CAZyme (155–306) (Table 2) and peptidase (332–404) encoding genes (Table 3) were found concordant with what is observed in other sequenced species of the *Mucoromycota* phylum as indicated in the MycoCosm protein database. In particular, the total number of CAZyme encoding genes in the *Mucor* genomes analysed in this study was lower than what is observed in average in *Ascomycota* and *Basidiomycota* (Additional file 3: Figure S3). Moreover, among CAZyme genes, those encoding glycosyl-transferases (GT) were found more numerous than those encoding glycoside hydrolases (GH) which is usual in the *Mucoromycota* but not in *Dikarya* (Additional file 3: Figure S3).

Clear differences in terms of CAZyme encoding gene composition among the different genomes were detected. Among the investigated species, *M. indicus* possessed the highest content in all the CAZyme gene classes. In contrast, the number of all classes of CAZyme genes (except those encoding redox enzymes classified in auxiliary activities -AA-) was drastically reduced in *M.*

**Table 2** Number of genes encoding Carbohydrate Active enZymes (CAZymes). Major CAZymes classes are shown separately, Auxiliary redox enzymes (AA), Carbohydrate-Binding Modules (CBM), Carbohydrate Esterases (CE), Glycoside Hydrolases (GH), Glycoside Transferases (GT) and Polysaccharide Lyases (PL). Enzymes substrats are indicated: Cell Wall (CW), Fungal Cell Wall (FCW) and Plant Cell Wall (PCW). The four isolates sequenced in the context of this study are in red

| Isolate | AA | CBM | CE | GH | GT | PL | Total CAZymes | CW | FCW | PCW |
|---|---|---|---|---|---|---|---|---|---|---|
| *M. endophyticus* CBS 385–95 | 12 | 2 | 15 | 56 | 67 | 2 | 155 | 10 | 19 | 8 |
| *M. fuscus* UBOCC-A-109160 | 12 | 7 | 18 | 68 | 107 | 1 | 214 | 12 | 21 | 12 |
| *M. lanceolatus* UBOCC-A-109153 | 10 | 5 | 16 | 62 | 105 | 2 | 201 | 11 | 19 | 12 |
| *M. racemosus* UBOCC-A-109155 | 11 | 5 | 20 | 81 | 115 | 2 | 235 | 18 | 22 | 14 |
| *M. griseocyanus* NBRC_6742 | 12 | 7 | 25 | 83 | 113 | 2 | 243 | 18 | 25 | 13 |
| *M. circinelloides* 1006PhL | 15 | 7 | 23 | 83 | 121 | 2 | 252 | 18 | 24 | 14 |
| *M. circinelloides* CDC-B8987 | 15 | 7 | 21 | 83 | 110 | 2 | 239 | 17 | 24 | 14 |
| *M. lusitanicus* CBS 277.49 | 11 | 5 | 23 | 75 | 112 | 2 | 229 | 16 | 22 | 12 |
| *M. velutinosus* CDC-B5328 | 16 | 6 | 24 | 78 | 112 | 2 | 239 | 17 | 22 | 13 |
| *M. indicus* CDC-B7402 | 25 | 11 | 34 | 96 | 136 | 3 | 306 | 20 | 28 | 15 |

**Table 3** Number of genes encoding proteolytic enzymes (peptidases) and their inhibitors (MEROPS database). The four isolates newly sequenced in the context of this study are in red

| Peptidases | | | | | | | |
|---|---|---|---|---|---|---|---|
| Isolate | Aspartic | Cysteine | Metallo | Serine | Threonine | Total peptidases | Inhibitors |
| *M. endophyticus* CBS 385–95 | 31 | 80 | 107 | 118 | 19 | 355 | 11 |
| *M. fuscus* UBOCC-A-109160 | 28 | 75 | 103 | 117 | 21 | 344 | 10 |
| *M. lanceolatus* UBOCC-A-109153 | 24 | 77 | 101 | 115 | 20 | 337 | 10 |
| *M. racemosus* UBOCC-A-109155 | 31 | 85 | 100 | 117 | 20 | 353 | 12 |
| *M. griseocyanus* NBRC_6742 | 32 | 84 | 101 | 110 | 20 | 347 | 10 |
| *M. circinelloides* 1006PhL | 32 | 82 | 101 | 114 | 20 | 349 | 13 |
| *M. circinelloides* CDC-B8987 | 30 | 78 | 94 | 109 | 21 | 332 | 11 |
| *M. lusitanicus* CBS 277.49 | 34 | 86 | 103 | 108 | 20 | 351 | 10 |
| *M. velutinosus* CDC-B5328 | 29 | 85 | 103 | 109 | 21 | 347 | 13 |
| *M. indicus* CDC-B7402 | 44 | 88 | 114 | 132 | 26 | 404 | 11 |

*endophyticus*. The number of genes coding for catabolic CAZymes was also reduced in *M. fuscus* and *M. lanceolatus*. Interestingly, these reductions were noticeable for genes coding for cell wall degradation enzymes and more specifically in plant cell wall (PCW) degradation enzymes in *M. endophyticus* (Table 2). Differences among groups of species sharing a same putative lifestyle in terms of number of CAZyme families encoding genes were illustrated by the principal component analysis (PCA) of CAZymes in particular along axes 1, 2 and 3 (Fig. 3). The first axis represented the opposition between the thermophilic opportunistic pathogen *M. indicus* and the endophyte *M. endophyticus*. This axis was constructed using mainly 28 CAZyme families (correlation above 0.75) all contracted in *M. endophyticus* compared to *M. indicus*, the other species presenting in almost all cases an intermediary pattern.

The second axis opposed *M. endophyticus* and *M. indicus* to all the other species. Three CAZyme families were mainly involved (correlation above 0.75) in this axis construction: GH8, GH20 and GT77. It is noteworthy that the GH20 β-hexosaminidase family was expanded in *M. endophyticus* and *M. indicus*, whereas the GH8 and GT77 tend to be reduced in the two latter species. The third axis represented the opposition between two species used for cheese ripening (*M. fuscus* and *M. lanceolatus*) and all the other species except *M. indicus*, the latter one presenting an intermediary pattern. Only four CAZymes families were mainly involved (correlation above 0.75) in this axis construction: GH9, GH37 and GT4 were contracted and GH45 expanded in *M. fuscus* and *M. lanceolatus* compared to almost all the other studied species.

Noteworthy, this analysis highlighted the GH3, GH5 (first axis) and GH9 (third axis) CAZyme families which are among the major cellulases involved in plant cell wall degradation. They were less represented in *M. endophyticus* and to a lesser extent in *M. fuscus* and *M. lanceolatus* (Additional file 4: Table S1).

Noteworthy, in addition to its expanded CAZyme repertoire, *M. indicus* also displayed the largest peptidase (Table 3) repertoire and, with *M. circinelloides* isolate 1006PhL, the highest frequency of genes coding for secreted proteins and small secreted proteins (SSPs) (Fig. 4). In contrast, *M. fuscus* and *M. lanceolatus* exhibited the lowest frequencies of secreted protein and SSP encoding genes (Fig. 4).

#### *Genes involved in secondary metabolism*

As secondary metabolism genes can be associated with habitat adaptation, genes encoding PKS, NRPS, TPS and DMATS were investigated. Among genes associated with terpene biosynthesis, some corresponding to squalene cyclases, squalene synthases, bifunctional lycopene cyclases, squalene/phytoene synthases and geranylgeranyl pyrophosphate (*ggpp*) synthases were identified in each species. The number of genes in each category was well conserved among the different species (Table 4).

In all studied species, a single complete *nrps* gene (i.e. having at least one condensation domain, one carrier domain, one phosphopantetheine attachment site and 1 AMP-binding domain) was detected. Three other genes containing three of the four mandatory NRPS domains were found in *M. indicus*, both of them lacking the condensation domain, and one in *M. lusitanicus*, lacking the AMP-binding domain. In all species, no gene coding for DMATS were identified.

Interestingly, some gene sequences encoding putative PKS can be determined by automated annotation; however, using a manual approach, these genes actually correspond to fatty acid synthases (FAS) type I encoding genes. They were detected in the investigated genomes: three in *M. indicus*, one in *M. fuscus* and *M. lanceolatus*

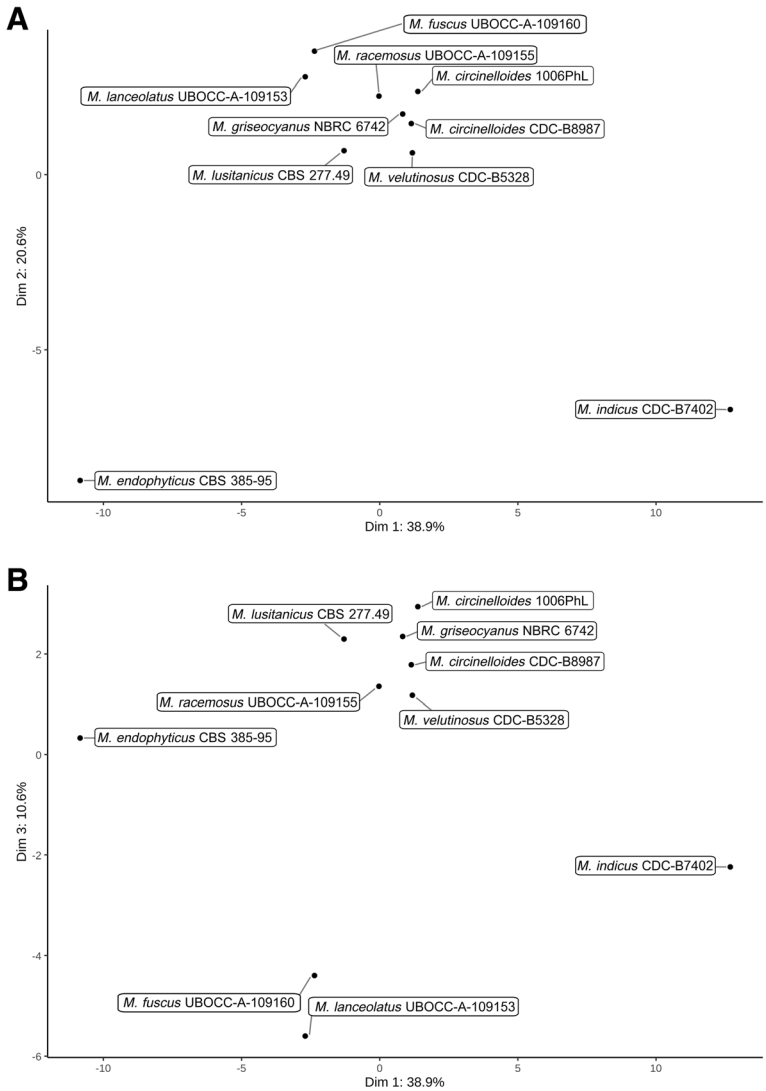


**Fig. 3** Representation of *Mucor* isolates repartition within the PCA analysis on their CAZYmes family proteins content. **a** Representation of the first two dimensions. **b** Representation of dimensions 1 and 3

and two in the other genomes. The different domains of these genes were similar in terms of composition and organization to FAS encountered in *Basidiomycota*, while FAS are encoded by two genes in *Ascomycota* (Fig. 5). In *M. indicus*, *M. racemosus* and in *M. velutinosus*, one of the FAS genes held a second KS domain. Recent transcriptomic analyses indicated that these genes were not expressed on PDA medium [23]. Noteworthy, none of the potential secondary metabolism associated genes determined in the different *Mucor* genomes were organized in metabolic clusters.

### *Iron uptake*

As the involvement of iron uptake in mucormycosis has been reported, the associated genes were studied [30]. Homologs of genes encoding proteins involved in the different iron uptake mechanisms identified so far in fungi [31] were found in the analysed *Mucor* genomes (Table 5; Fig. 6).

Regarding the siderophore-mediated iron uptake, genes playing a role in the carboxylate and hydroxamate siderophore synthesis were searched for. At least one ortholog of the *R. delemar rfs* gene, necessary for the carboxylate siderophore rhizoferrin production [33] was found in each *Mucor* species. Other genes that might be involved in this rhizoferrin-mediated iron uptake mechanism were identified based on their homology to the bacterial genes of the *Francisella tularensis* rhizoferrin operon [34]. Homologs of the *FslB* and *FslC F. tularensis* genes were detected in each *Mucor* genome and numerous potential *Mucor*

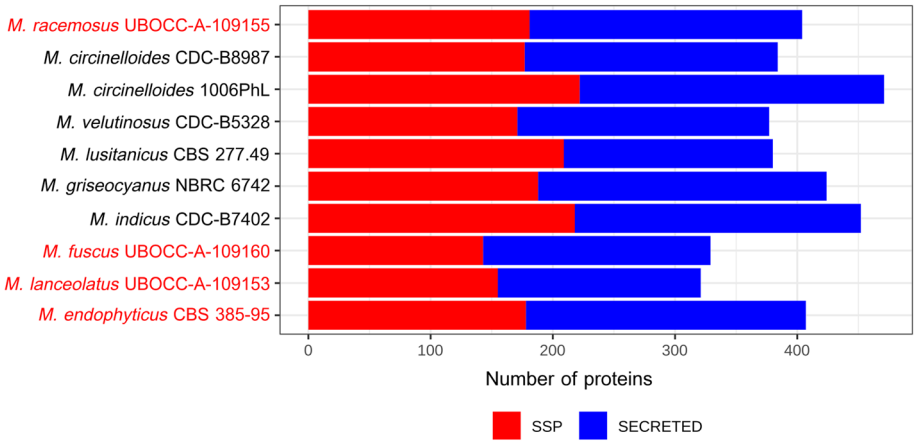


**Fig. 4** Representation of the number of secreted proteins within the genomes of the studied *Mucor* isolates. SSP: secreted proteins containing less than 300 aa. Secreted: secreted proteins with more than 300 aa. The four isolates sequenced in the context of this study are in red

genes belonging the major facilitator family matched to *FslD*. However, *FslA*, *FslE*, *FslF* and the operon regulator *Fur* could not be detected in *Mucor* genomes. Genes involved in hydroxamate siderophore synthesis were not detected but predicted orthologous genes corresponding to the *Aspergillus* MirB siderophore permease (group 1) encoding gene and another gene coding for a MirB-like siderophore permease (group 2) were identified (Table 5).

Regarding the reductive iron assimilation iron uptake, homologous sequences of the gene encoding the *FTR1* high-affinity permease and *fet3* ferroxidase genes were detected (*fet3a* was not detected in *M. fuscus*, *M. lanceolatus* and *M. racemosus* genomes). Except for the FTR1 encoding gene, no gene involved in heme uptake was identified. The FET4 low affinity iron permease encoding gene was identified in the different *Mucor* genomes. When focusing on the iron uptake regulation, homologs to the *SreA* iron uptake

**Table 4** Number of genes involved in secondary metabolites found in the 10 studied *Mucor* isolates. The four species sequenced in this study are in red. For each gene category, maxima are highlighted in orange and minima in blue

| | **Isolate** | *M. endophyticus* CBS 385-95 | *M. fuscus* UBOCC-A-109160 | *M. lanceolatus* UBOCC-A-109153 | *M. racemosus* UBOCC-A-109155 | *M. griseocyanus* NBRC 6742 | *M. circinelloides* 1006PhL | *M. circinelloides* CDC-B8987 | *M. lusitanicus* CBS 277.49 | *M. velutinosus* CDC-B5328 | *M. indicus* CDC-B7402 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Terpenes | Squalene cyclase | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Squalene synthase | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| | Lycopene cyclase & squalene/phytoene synthase* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Geranylgeranyl pyrophosphate synthase | 4 | 5 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 4 |
| NRPS | NRPS | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | NRPS-like 3/4 domains | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| PKS | PKSI | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| | 3-oxoacyl synthase II (PKS-like) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*These genes encode a putative bifunctional enzyme

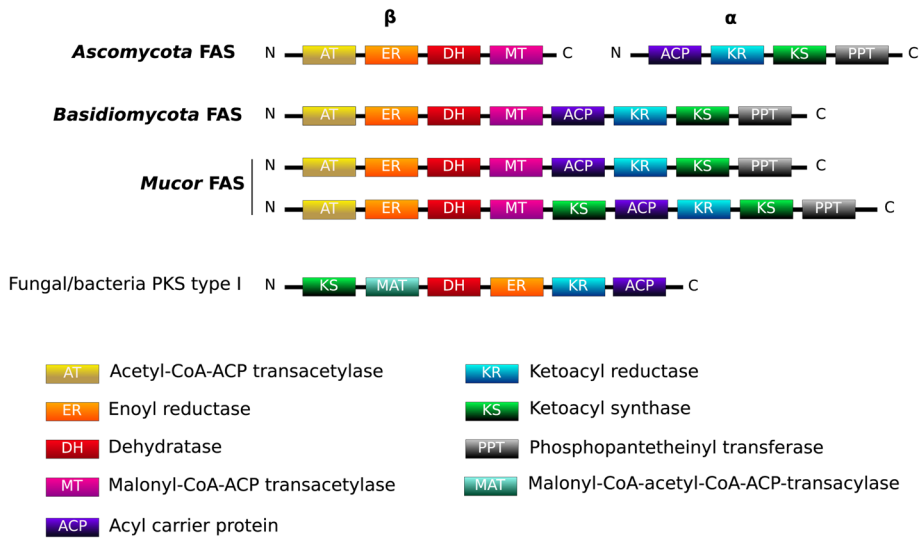


**Fig. 5** Domain organization of FAS within the studied *Mucor* genomes in comparison to *Ascomycota* and *Basidiomycota* reported FAS organizations

repressor gene were detected but no gene involved in activation of iron acquisition pathways such as *HapX* or *Aft* genes were identified. Two ferritin encoding genes of similar lengths were present in each studied species; in each case, the two genes shared 70% similarity and 40% identity. Interestingly, the number of genes involved in iron uptake was always higher in the genome of the opportunistic pathogen *M. indicus* than in the other genomes (Table 5). Indeed, in the *M. indicus* genome, (i) for the siderophore pathways, the *rfs* rhizoferrin biosynthesis gene was duplicated as well as a MirB-like siderophore permease encoding gene; (ii) for the reductive iron assimilation (RIA), the *fet3b* and *fet3c* ferroxidase genes were duplicated; (iii) for heme degradation, a supplementary heme oxygenase gene homolog was detected and (iv) for direct $Fe^{2+}$ uptake, three orthologs of the FET4 low affinity permease were identified, whereas two genes were found for the *M. circinelloides* complex and one for the other studied species. Finally, only one copy of the *sreA*-like gene, involved in down regulation of iron acquisition, was found in *M. indicus* whereas at least two copies were found for all other species.

On the contrary, the two cheese-associated species (*M. fuscus* and *M. lanceolatus*) exhibited a reduced number of genes involved in iron uptake within their genomes. *MirB*-like siderophore permease encoding gene was absent from the *M. fuscus* genome whereas one copy of *fslB*-like siderophore permease gene was lost. *M. lanceolatus* lost one copy of the cell surface receptor *fob* gene and of the *FTR1* high affinity permease, both species lacked the *fet3a* ferroxidase genes. The latter genes were also absent from the cheese contaminant *M. racemosus* genome.

### *Antifungal resistance*

According to their habitat, fungi can be confronted to chemical biocide. In particular, at the hospital or in agriculture, fungicides are used and fungi can develop resistance. Based on the genome comparison we focused our attention on the antifungal resistance of the studied *Mucor* to triazoles.

The deduced aminoacyl sequences of the two lanosterol 14α-demethylase paralogues CYP51 F1 and CYP51 F5 involved in ergosterol biosynthesis were globally well conserved over the *Mucor* genomes with the occurrence of the F128 residue in Helix I of CYP51 F5 which has been reported to be responsible for short-chain azole resistance in *Mucorales* [35]. However, a two-residue substitution (AA to TS at residues 290–291) of Helix I of CYP51 F5 in *M. lanceolatus* was observed (Fig. 7).

According to Caramalho et al., this residue could play a role in the intrinsic resistance against triazoles. These observations raised the question of the impact of these amino acid substitutions on *M. lanceolatus* azole susceptibility. Growth studies on azole supplemented media clearly showed that *M. lanceolatus* was the most susceptible species to both long- (posaconazole) and short-chain (voriconazole) azoles among the tested *Mucor* species (Table 6) [35].

## Discussion

Although benefiting from a growing interest due to their involvement in mucormycosis but also their biotechnological potential, a restricted number of *Mucor* genomes is available to this date, leading to scarce whole genome comparative studies [13, 15, 36]. These comparative

**Table 5** Number of genes involved in iron uptake found in the 10 studied *Mucor*. The four species sequenced in this study are in red. For each gene category, maxima are highlighted in orange and minima in blue. Proteins encoded by the different genes and their role in iron uptake mechanisms are presented in Fig. 6

| | | **Isolate** | *M. endophyticus* CBS 385-95 | *M. fuscus* UBOCC-A-109160 | *M. lanceolatus* UBOCC-A-109153 | *M. racemosus* UBOCC-A-109155 | *M. griseocyanus* NBRC 6742 | *M. circinelloides* 1006PhL | *M. circinelloides* CDC-B8987 | *M. lusitanicus* CBS 277.49 | *M. velutinosus* CDC-B5328 | *M. indicus* CDC-B7402 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| uptake | Siderophore (sid) mediated iron uptake | Sid-transporter: mirB-like group 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | Sid-transporter: mirB-like group 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| | | Sid-transporter: fslB | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | | Sid-biosynthesis: rfs | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| | | Sid-biosynthesis: fslC | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Reductive Iron Assimilation (RIA) | High-affinity iron permease: FTR1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| | | Ferroxidase: fet3a | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | Ferroxidase: fet3b | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| | | Ferroxidase: fet3c | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| | | other Ferric reductase: Fre | 5 | 3 | 3 | 4 | 4 | 4 | 4 | 3 | 4 | 4 |
| | | Ferrioxamine binding: Fob | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | Heme degradation | Heme oxygenase: HOXG | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| | Low affinity iron permease | Low affinity iron permease: FET4 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 |
| Iron regulation | | Transcription factor: sreA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| Iron utilization | | Iron-binding protein frataxin: yfh1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Iron storage | | Vacuolar iron importer: ccc1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | | Vacuolar iron transporter: smf3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 |
| | | Ferritin | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

studies mainly focused on human/clinical environments species. Yet, only a handful of *Mucor* species are known to cause human infections [2]. In this context, this study expanded the range of *Mucor* genomes available by including genomes from species collected from non-clinical environments. In particular, *M. endophyticus* was only reported as a wheat endophyte [27] and *M. lanceolatus* was, so far, only collected from cheeses [26].

The determined phylogeny, based on a large set of orthologous genes (51), that integrated the four genome sequences obtained in this study is partially concordant with published phylogenies. In particular it confirmed the phylogenetic relationships amongst different members of the *M. circinelloides* complex [4] but was non-concordant with the latest published phylogeny concerning the phylogenetic position of *R. microsporus* CDC-B9738 and CDC-B9645 isolates, as they appeared in a sister group of the *M. circinelloides* complex and closely related to *M. racemosus*. This was observed both in the present study and that of Chibucos et al. (76 analysed orthologs) while they clustered into the *R. microsporus* clade with *M. circinelloides* species as an outgroup in Gryganskyi et al. study (192 orthologues analysed) [13, 25]. Topologies may vary depending on the selected genes and on the reconstruction pipeline. The contrasted placements of the 2 *R. microsporus* isolates, which harbour larger-sized genomes (Fig. 2c), among the different studies may arise from the whole/partial-genome duplication events and/or hybridization observed in their genomes [25]. As a confirmation, when considering gene families involved in secondary metabolism and iron acquisition (which were investigated in this study by comparative genomics approaches), *R. microsporus*

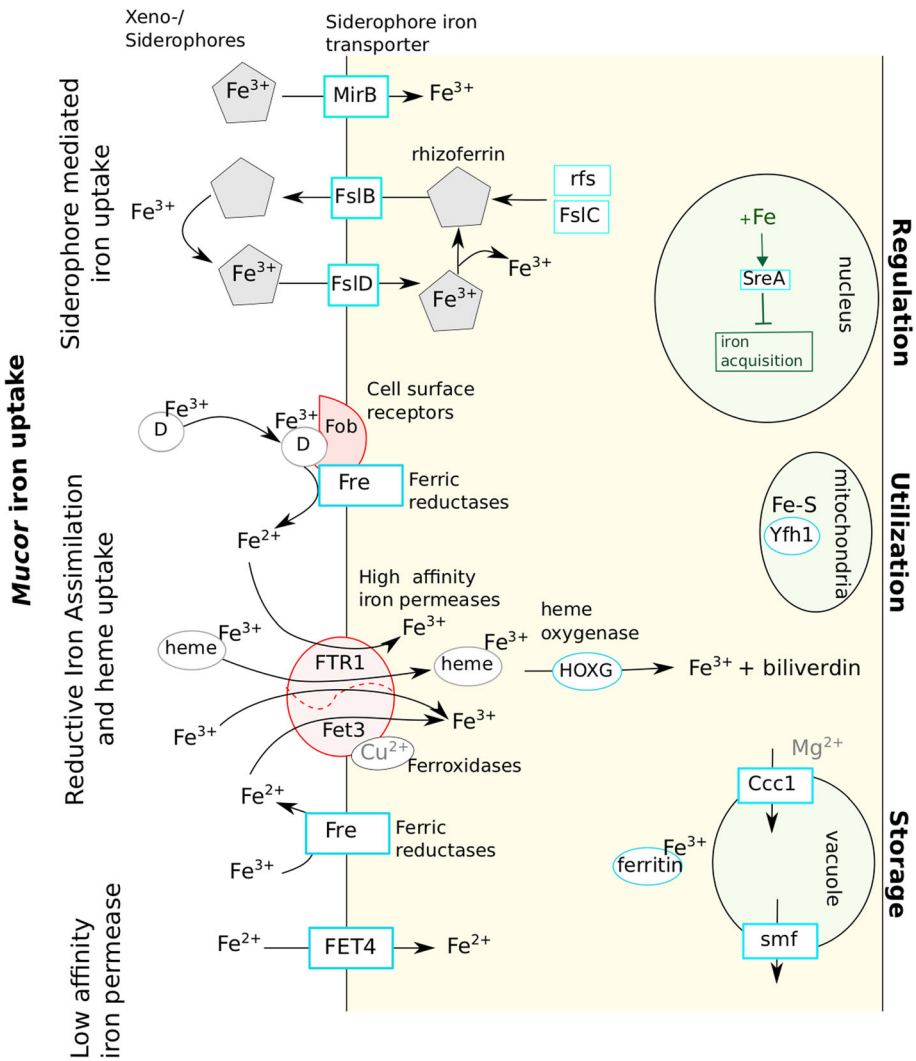


**Fig. 6** Proposed mechanisms of iron uptake by *Mucor* spp. as well as elements used for iron regulation, utilization and storage. Elements in red were described in literature as important for *Mucor* pathogenicity [13, 16, 32]. D: Deferoxamine

CDC-B9738 and CDC-B9645 genes were often found duplicated. Phylogenetic reconstructions performed using the duplicates yielded incongruent topologies with either CDC-B9738 or CDC-B9645 isolates clustering with *R. delemar* clade or alternatively with *M. racemosus* clade suggesting a possible hybrid genomic content for these two isolates.

Whatever the phylogenetic placement of the studied species or their proposed habitat/lifestyle, the current study revealed that the gene features among *Mucor* species (gene number, size, exon length etc.) were globally conserved. However, as indicated by the lack of macro- and micro-synteny (data not shown), species within this genus experienced extensive genomic rearrangements. These rearrangements can be partially explained by transposable elements (TE), which displayed high degree of diversity within the available genomes and have already been reported to have a major role in fungal genomic diversity and genome evolution [37].

The average genome size of the *Mucor* species analysed in this study (39 Mb) is congruent with what is observed at the scale of the *Mucoromycotina* subphylum (38 Mb) and also in *Ascomycota* (37 Mb). The gene number (9997 to 12,571) is also in concordance with the gene numbers observed in *Mucoromycotina* and *Ascomycota* [38]. The core genome detected in this study includes approximately 6000 genes. The species-specific genes and the genes shared by species with similar lifestyles (data not shown) mostly encode proteins with unknown functions. They are thought to determine specific traits, such as adaptation to different environmental niches or preferential colonization of certain habitats, and determining their function would be of special interest.

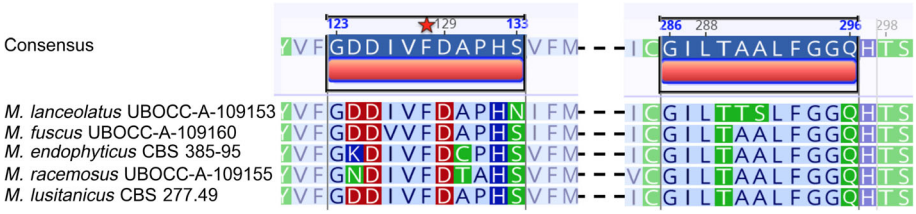


**Fig. 7** Partial amino acid sequence alignment of lanosterol 14α-demethylase F5 in different *Mucor* isolates. The consensus sequence is represented in blue and polar amino acids appear in green. ML: *M. lanceolatus;* MF: *M. fuscus;* ME: *M. endophyticus;* MR: *M. racemosus;* MLu: *M. lusitanicus*

As osmotrophic organisms, fungi depend, for gaining nutrients, on lytic enzyme secretion to externally break down surrounding resources prior to absorption of the catabolized products [39]. Depending on the lifestyle (saprotrophic, pathogenic or mutualistic) and the habitat, the way of gaining nutrients and energy may vary. In particular qualitative and quantitative differences in terms of CAZymes can be potent reporters of fungal lifestyle [38, 40–42]. In the present study, diverging CAZyme repertoires among species or species groups sharing the same ecology were identified. The genome of *M. endophyticus* CBS385–95, isolated from wheat leaves and described as an endophyte, shares similar traits with genomes of other fungal endophytes. It harbours a low number of catabolic CAZymes (CBM, GH and PL) and esterase (CE) genes [43] including genes coding for plant cell wall degrading enzymes, such as the GH5, GH3 and GH9 fungal cellulases encoding genes. This observation suggests that *M. endophyticus*, like other fungal endophytes and ectomycorrhizal fungi, is indeed adapted to intercellular growth and to the avoidance of the plant immune system [43, 44]. The two species almost only isolated from cheeses and involved in cheese ripening (*M. fuscus* and *M. lanceolatus*) also displayed, to a lesser extent than *M. endophyticus*, a reduced repertoire of carbohydrate active enzyme encoding genes. The reduction was less significant for the GT encoding genes whose roles are not directly connected to the external environment. It could be hypothesized that this reduced set of catabolic CAZymes is an imprint of specialization to a less ubiquitous lifestyle and an adaptation to the cheese environment. In contrast, the thermophilic opportunistic pathogen *M. indicus* harbours the largest set of CAZyme genes. This trait is not shared by *M. circinelloides*, a species that may cause mucormycosis. This could be related rather to the high capacity of *M. indicus* to degrade plant materials [45] than to an opportunistic pathogenic lifestyle. However, it is worth noting that *M. indicus* possesses a much larger protease repertoire compared to the other *Mucor* species including the *M. circinelloides*. Given that proteases [46, 47] were found clearly correlated to pathogenicity in *Ascomycota* opportunistic pathogens one might expect a higher pathogenicity potential of *M. indicus* over *M. circinelloides*.

Examination of secondary metabolic pathways can also shed light on fungal ecology and adaptation to a particular lifestyle [48]. In this study, we investigated more specifically genes that are essential to most of the secondary metabolite biosynthesis (PKS, NRPS, DMATS …) but also genes involved in iron uptake which have been shown to be fundamental for fungal virulence in opportunistic pathogens [49, 50] and for cheese colonizing microorganisms given that cheese is a strongly iron-restricted medium [51]. The search for genes involved in secondary metabolite biosynthesis are often initiated by

**Table 6** Growth diameter (cm) of representative *Mucor* isolates on RPMI medium supplemented or not with different concentrations of voriconazole (VCZ) or posaconazole (PCZ)

| | VCZ concentration | | | | | | PCZ concentration | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Isolate | 0 mg/L | 10 mg/L | 50 mg/L | 100 mg/L | 500 mg/L | 1000 mg/L | 0 mg/L | 10 mg/L | 50 mg/L | 100 mg/L | 500 mg/L | 1000 mg/L |
| *M. endophyticus* CBS 385-95 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | 0 | 0 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | 0 | 0 |
| *M. fuscus* UBOCC-A-109160 | ≥ 5,4 | 3,2 | 0,8 | 0,2 | 0,2 | 0,2 | ≥ 5,4 | 3 | 3 | 3 | 1,5 | 0,2 |
| *M. lanceolatus* UBOCC-A-109153 | ≥ 5,4 | 4,5 | 0 | 0 | 0 | 0 | ≥ 5,4 | 0 | 0 | 0 | 0 | 0 |
| *M. racemosus* UBOCC-A-109155 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | 3,2 | 0,5 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 |
| *M. lusitanicus* CBS 277.49 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | 0,3 | 0,3 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 | ≥ 5,4 |

the search for PKS, NRPS or DMATS genes clustered with additional genes involved in the considered biosynthesis pathway and known as BGC (biosynthetic gene cluster). In the framework of this study, no typical BGCs involved in secondary metabolism, including a PKS, FAS, NRPS, NRPS-like gene or terpene synthesis related genes, were detected within the studied *Mucor* genomes. Still, secondary metabolites pathways have been characterized in *Mucorales* [3] and different genes involved in secondary metabolism were identified in the present and previous studies. The apparent absence or at least scarcity of BGCs within the genus *Mucor* (and possibly at a broader scale within the *Mucoromycotina*) might appear enigmatic since the genus encompasses species with contrasting lifestyles similar to what exists in *Ascomycota* in which BGCs are abundant. This raises the question why the apparent selective advantage conferred by BGCs to higher fungi would not apply to *Mucor* species. The answer to this question might lie in some ecological specificities of the *Mucorales* which are considered as ruderal species avoiding stress and competition [52, 53] or to structural specificities such as their coenocytic structure [54], and may shed new light regarding the BGC evolution in eukaryotes. All *Mucor* genomes investigated in this study included a NRPS gene which role has still to be determined as well as different genes encoding for enzymes associated with terpene biosynthesis. As noticed by Voigt et al. who analysed different *Mucor* genomes, no PKS genes but genes with a typical structure and domain order of Type I fatty acid synthases (FAS) were detected in the different *Mucor* genomes investigated in the present study [3]. These *Mucor* FAS have their domains allocated within a single gene as in *Basidiomycota* and not over two genes as in *Ascomycota* (Fig. 5) [55, 56]. This result supports the hypothesis of a primordial contiguous FAS gene encoding the entire FAS [55]. Most of the *Mucor* genomes analysed here include two putative FAS genes, but a single FAS gene was found in the cheese-related species *M. lanceolatus* and *M. fuscus*, and up to three genes were found in the opportunistic pathogen *M. indicus*. Whatever the number of FAS genes included in the genome, only one FAS was found expressed when the isolate was cultivated on PDA medium [23]. It could be hypothesized that the additional FAS genes have lost their functionality or display a different role since FAS gene are expected to be expressed in such conditions to participate to the fatty acid synthesis for membrane synthesis and energy storage [56]. Whether or not this additional FAS might be involved in secondary metabolisms like specialized secondary metabolism FAS [57] has to be determined.

Among the secondary metabolites, terpenes play a role in natural pigment synthesis such as the carotenoid biosynthesis [3]. Carotenoid synthesis has been described in *M. circinelloides* and in particular in overproducing isolates [3, 58, 59]. Although the different species within the *Mucoromycotina* subphylum are expected to produce and accumulate large amount of carotenoids, differences might exist among species [60]. Terpenes play also an important role in flavour production which is an important trait for cheese-ripening fungi that could have been under human-directed selection in species used for cheese making [61]. This study indicated that the main genes involved in terpene biosynthesis were conserved among the analysed *Mucor* species without any important differences in terms of gene number with the exception of GGPP genes for which the number is lower in the endophytic species as well as in *M. griseocyanus* and in the *M. indicus* opportunistic pathogen.

Iron is an essential nutrient involved in variety of cellular processes such as respiration, oxidative stress and amino acid synthesis [62]. Iron uptake has been described as a virulence factor for pathogenic fungal isolates [49, 50] and of primary importance in cheese microorganisms since cheese is a highly iron-depleted medium [51]. In fungi, four different iron uptake mechanisms have been described so far: (i) siderophore mediated $Fe^{3+}$ uptake, (ii) reductive iron assimilation (RIA), (iii) heme uptake and (iv) direct $Fe^{2+}$ uptake [49] (Fig. 6). Homologs of genes coding for proteins participating to the four mechanisms were found in the analysed *Mucor* genomes. These results suggest that the different *Mucor* species investigated here rely on carboxylate siderophore rhizoferrin as it is the case for *M. circinelloides* CBS 277.49 and *R. delemar* [33]. This type of siderophore is not used by *Ascomycota* and *Basidiomycota*, and would thus be a specificity of early diverging fungi in the fungal kingdom. Noteworthy, rhizoferrin encoding gene sequences have only been described in bacterial genomes so far [32, 63] and coding sequences pertaining to the rhizoferrin operon in *F. tularensis* were used here to search for homologs in *Mucor* spp. In *F. tularensis*, genes involved in rhizoferrin synthesis and transport are located in an operon regulated by the *Fur* gene [34]. Based on these results, new candidate genes involved in rhizoferrin synthesis, import and export are proposed in the present study (Fig. 6). Although no actual evidence of *Mucor* ability to produce hydroxamate siderophores was detected, some of these siderophores could be used by *Mucor* species as xeno-siderophores as suggested by the presence of *mirB*-like siderophore transporter genes in some of the *Mucor* genomes. It is worth noting that the bacterial siderophore deferoxamine is also used by *Mucor* spp. as xeno-siderophores [64].

Interestingly, the three isolates sampled from cheese presented a reduced number of genes related to iron acquisition compared to the other isolates. Indeed, the genomes of these isolates lack a FTR1 gene copy as well as

the *fet3a* gene. It could be hypothesized that the latter genes would have a specific role in *Mucor* pathogenicity since *R. delemar* mutants with FTR1 reduced gene copies or with decreased FTR1 expression had reduced virulence in a deferoxamine-treated mouse model of mucormycosis [30]. The *fet3a* gene appeared as the less important among the *fet3* genes regarding *M. circinelloides* pathogenicity, but inactivation of two *fet3* genes led to a reduced virulence [17] and the loss of *fet3a* led to an increased sensibility to the mutations on *fet3b/fet3c* in terms of fungal pathogenicity. Furthermore, one copy of the ferroxamine binding (Fob) cell surface protein gene is absent in the *M. lanceolatus* genome, the production of this protein being required for full virulence of *R. delemar* in a deferoxamine-treated mouse model of mucormycosis [14]. *M. fuscus* also lacks *MirB*-like and one copy of *fslB*-like siderophore permease genes, which might reduce its potential to acquire iron. These results provide possible evidence of safety of these food-related species although the number of studied isolates is too low to definitely assert it. On the contrary, *M. indicus*, a species that is considered as the most threatening opportunistic human and animal pathogen amongst the *Mucor* species [7] harbours an increased set of genes involved in iron uptake which might be an asset to its opportunistic pathogenic lifestyle.

Susceptibility to antifungal drugs, and in particular to azole antifungal agents which are widely used for mucormycosis treatments as well in agriculture [35, 65], is of interest as it may vary in *Mucor* species according to their specific habitats, e. g. cheese isolates are probably less exposed to antifungals and antifungal resistance is unlikely to offer a selective advantage to non-pathogens. Azole drugs inhibit lanosterol 14α-demethylase (LDM) thus blocking ergosterol biosynthesis and resulting in a build-up of toxic sterols [66]. In *Mucorales*, two lanosterol 14α-demethylase paralogues CYP51 F1 and F5 coexist with a substitution in CYP51 F5 at residue 128 responsible for innate resistance against short-tailed triazoles, and a V to A substitution at residue 291 of CYP51 F5 also potentially playing a role [35]. Interestingly, we found that *M. lanceolatus*, only isolated so far from cheese, did not harbour the V to A substitution at residue 291 of the CYP51 F5 predicted protein but bears instead a unique two aminoacyl substitution (AA to TS) at residues 290 and 291 (in consensus sequence). Azole tests performed in this study demonstrated that these substitutions were associated with a notable higher susceptibility to both short- and long-tailed azoles.

## Conclusions

In conclusion, this study expanded the range of *Mucor* genomes available by including genomes from species with contrasted lifestyles represented by isolates collected from non-clinical environments (more specifically, cheese and plant). In addition to yield a better overview of the *Mucor* pan-genome, the expanded range of genomes allowed identifying contrasting features that could contribute to habitat and niche adaptation although distinguishing divergences due respectively to evolutive adaptation or to non-ecologically based evolutionary forces may appear difficult given that the different taxa did not diverged recently. The obtained data will allow further investigating the link between genetic and biological data, especially in terms of niche or habitat adaptation.

## Materials and methods

### Biological material

The genomes of 12 representative isolates were investigated in the present study (Table 7). Four of them were sequenced in the framework of this study while the eight others were publicly available [12, 13, 28]. *M. fuscus* UBOCC-A-109160, *M. lanceolatus* UBOCC-A-109153, *M. endophyticus* CBS 385–95 (UBOCC-A-113049) and *M. racemosus* UBOCC-A-109155 used for genome sequencing were cultivated in the dark at 25 °C on PDA solid medium (Difco Laboratories, Detroit, Michigan). Spore suspensions of each isolate were produced as previously described by Morin-Sardin et al. [7]. Concentrations were adjusted to $10^7$ to $10^8$ spores·$mL^{-1}$ prior to storage at − 80 °C until use. For genomic DNA extraction, each of the studied monospore isolates was grown on PDA solid medium at 25 °C for 7 days.

### Species assignment within *M. circinelloides* complex

Species assignment of isolates pertaining to the *M. circinelloides* complex was performed using a Maximum Likelihood phylogenetic reconstruction based on a *cyclopropane-fatty-acylphospholipidsynthase* gene (*cfs*) sequence alignment including *cfs* sequences available in Wagner et al. [4].

### Genome sequencing and assembly

Genomic DNA from *M. fuscus, M. racemosus, M. lanceolatus* and *M. endophyticus* was extracted from fresh mycelium, following the genomic DNA extraction method as outlined by the 1000 Fungal Genomes Project (http://1000.fungalgenomes.org/home/protocols/high-quality-genomic-dna-extraction/#Fulton1995) which is based on a protocol by Spanu et al. [72]. with an optional step using Qiagen genome-tips (Qiagen). Due to the low efficiency of the CTAB method for *M. lanceolatus*, genomic DNA of this isolate was also extracted following the protocol developed by Cheeseman et al. with a purification by a cesium chloride gradient with DAPI [73].

Genomes were sequenced with Illumina technology (San Diego, CA) at different sequencing facilities

**Table 7** List of isolates used in this study and their reported habitat. The four isolates sequenced in the context of this study are in red

| Species | Isolate | Isolation source | Reported habitat of the species | Reported habitat references | Genome reference or accession |
|---|---|---|---|---|---|
| *M. endophyticus* | CBS 385–95 (UBOCC-A-113049) | *Triticum aestivum*, leaves | Plant endophyte | [27] | This study |
| *M. fuscus* | UBOCC-A-109160 | Cheese | Cheese | [26] | This study |
| *M. lanceolatus* | UBOCC-A-109153 | Cheese | Cheese | [26] | This study |
| *M. racemosus* | UBOCC-A-109155 | Cheese | Cheese, yogurt, walnuts, sausages, grassland soil, decaying vegetables, human | [24, 26, 67] | This study |
| *M. griseocyanus*[a] | NBRC 6742 | Unknown | Vegetable, tanned sole leather | [68] | BBKB00000000 |
| *M. circinelloides* f. *circinelloides* | 1006PhL | Skin of a healthy human | Cheese, sufu starter, decaying vegetables, human, soda, air, soil, dung, sediment | [22, 24, 26, 69] | [29] |
| *M. circinelloides* | *CDC-B8987* | Human BL line | | | [13] |
| *M. lusitanicus*[b] | CBS 277.49 (UBOCC-A-108085) | Unknown | | | [12] |
| *M. velutinosus*[c] | *CDC-B5328* | Human: nasal | | | [13] |
| *M. indicus* | *CDC-B7402* | Human: unknown | Human, dung, *Dioscorea tuber*, sorghum malt | [24, 70, 71] | [13] |
| *R. microsporus (formerly M. racemosus*[c]*)* | *CDC-B9645* | Clean room floor | Human, dust, sorghum malt, stored cereals | [13, 24] | [13] |
| *R. microsporus (formerly M. racemosus*[c]*)* | *CDC-B9738* | Human abdomen | | | GCA_000697275.1 |

*CBS (Centraalbureau voor Schimmelcultures, Netherlands), NBRC (Biological Ressource Center, NITE), NRRL (Agricultural Research Service Culture Collection, USA), CDC (Centers for disease Control and Prevention) and UBOCC (University of Brest Culture Collection)*
[a]Originally *M. ambiguus* NBRC 6742. [b]Originally *M. circinelloides* CBS 277.49. These isolates were reassigned according Wagner et al. [4]. [c]Initially referenced as *Mucor racemosus* [13] but later assigned to *Rhizopus microsporus* [25]

(Additional file 5: Table S2). For each of the four species, DNA were paired end sequenced (read length 2 × 100 bp, insert size 500 bp). An additional mate pair sequencing was performed for *M. lanceolatus* and *M. endophyticus* (read length 2 × 100 bp, insert size 9–12 kb). Sequences were quality checked with FastQC [74]. Adaptors were removed, reads were quality trimmed (bases kept had a phred score above 25) and reads shorter than 20 bp were dropped with Cutadapt [75]. Mate pair reads of *M. lanceolatus* were mapped with STAR [76] on a preliminary version of the assembly (by providing only *M. lanceolatus* paired end data to CLC Genomics Workbench -CLCbio, Seoul, Korea-). Mate pair reads separated by less than 500 bp and oriented in forward-reverse were dropped. This new set of reads was used in further *M. lanceolatus* assemblies. *M. lanceolatus* and *M. endophyticus* genomes were assembled using Velvet [77] (option "shortMatePaired", k-mer of 67 for *M. lanceolatus* and k-mer of 85 for *M. endophyticus*), while *M. racemosus* and *M. fuscus* genomes were assembled with SOAPdenovo [78]. Genome assembly quality was checked with BUSCO v3 [79] using the fungal dataset and *Rhizopus* Augustus training.

### Genome annotation of the four newly sequenced genomes

Genome assembly scaffolds were annotated using combinations of ab initio predictors, RNAseq data support and homology search. As for ab initio predictors, Genemark-ES [80], with self-training, and Augustus [81], with *Rhizopus* training available within the Augustus tool, were used. RNAseq transcripts were extracted and sequenced as previously described in Lebreton et al. and reconstructed using two methods [23]: (i) by mapping RNAseq reads on genome with STAR [82] and reconstructing transcripts with Cufflinks [83], and (ii) by de novo transcript reconstruction with Trinity [84] and mapping the obtained transcripts on the genomes with GMAP [85]. Predicted proteins of *M. lusitanicus* [12], *R. delemar* [9] and *P. blakesleeanus* [12] were searched on genomes with Exonerate [86]. Consensus gene models were generated from all predictions by EVidenceModeler [87].

The obtained gene predictions were functionally annotated as follows: transmembrane domains were predicted with TMHMM [88], peptide signal with SignalP v4 [89] and Pfam domains with HMMER [90] using the PFAM-A database [91]. Sequence homologies were searched

using tBLASTx and BLASTp [92] (with an e-value threshold inferior to $10^{-5}$, against Swissprot-Uniprot and Uniref90 databases as well as *M. lusitanicus* CBS 277–49, *R. delemar* RA-99880 and *P. blakesleeanus* NRRL1555 filtered proteins obtained from the JGI platform [93]. EC numbers were predicted using PRIAM [94] and were transferred from homology researches. GO terms were transferred from homology search. Gene names were assigned with AHRD (Automated Assignment of Human Readable Descriptions) available on Github (https://github.com/groupschoof/AHRD).

Non-coding RNA were predicted with tRNAscan-SE [95], RNAmmer [96] and Infernal [97] using the Rfam database [98]. The obtained data were integrated in an instance of the genome viewer Apollo [99] allowing experts to validate gene prediction quality and perform manual curation.

### Complementary annotation of the full set of genomes

Transposable elements (TE) were annotated using the REPET pipeline [100] that includes a de novo prediction and TE classification [101]. Carbohydrate-active enzymes were searched using dbCAN2 [102] based on sequences available in the CAZy database [103] with HMMER (E-Value $<1e^{-15}$, coverage > 0.35), DIAMOND [104] (E-Value $<1e^{-102}$) and Hotpep [105] (Frequency > 6.0, Hits > 2.6). Only annotations predicted by at least two different tools were subsequently considered. Peptidases and their inhibitors available in the MEROPS database [106] were searched using BLASTp (E-Value $10^{-5}$).

### Phylogenetic reconstruction

Predicted proteomes of the 12 studied isolates, as well as those of *R. delemar* and *P. blakesleeanus* (both latter species being considered as outgroups), were compared based on sequence similarity to identify orthologous proteins using the Orthofinder v.2.2.0 software [107] (E-value $10^{-5}$, inflation 1.5). The 64 obtained single copy orthologs were selected to reconstruct the phylogeny of the studied species. Multiple alignment was inferred using PRANK v.170427 [108], run with default settings. Spuriously aligned regions were excluded with TrimAl v1.4.r15 [109] with a 0.2 gap threshold. Based on the alignments, 13 orthogroups were manually discarded due to low percentage of identical sites or high number of gaps among orthologs. Subsequent alignments were concatenated in a supermatrix of 23,398 sites. This matrix was used to reconstruct the species tree in one hand using Bayesian Monte Carlo Markov Chain (MCMC) samples with PhyloBayes v3.3 MCMC samplers [110] with a CAT+GTR model and 3 chains and in the other hand by Maximum Likelihood inference with RAxML PTHREADS v. 8.2.9 [111], with a partitioned LG + G model, in which each data partition represented a single input gene family. A bootstrap analysis with 100 replicates under the same model was performed in RAxML in order to assess tree branch support. To confirm the obtained phylogenetic position of the 2 *R. microsporus* isolates, an expanded species tree including 11 other sequenced *Mucoromycotina* (25 isolates overall) was reconstructed as follows. The 29 single copy genes families containing at least 20 isolates were independently aligned with PRANK v.170427 and trimmed with TrimAl v1.4.r15. For each gene family a gene tree was reconstructed using RAxML with the auto-detection of the best amino-acid model and a bootstrap analysis of 100 samples. The species tree was then reconstructed using the 29 gene trees with Clann v4.2.2 using a bootstrap analysis of 100 samples [112]. The obtained RAxML tree of the main species was used to estimate the divergence time between species with the Langley-Fitch method with r8s v1.8 [113] by calibrating against the assessed origins of *P. blakesleeanus* and *R. delemar* at 468 MY [114].

### Evolution of genes families

Based on OrthoFinder results and the obtained ultrametric tree, expansion and contraction of gene families were reconstructed with CAFE v4 [115]. Birth and death parameters were estimated independently using orthologous groups containing less than 75 genes per isolate. The analysis was done on all isolates except *R. microsporus* CDC-B9645 and *R. microsporus* CDC-B9738. Rapidly evolving families were predicted by CAFE using the Viterbi algorithm.

### Secondary metabolism related genes

Secondary metabolism associated genes (polyketide synthase -PKS-, non-ribosomal peptide synthetase -NRPS-, terpene synthase -TPS-, dimethylallyl tryptophan synthase -DMATS-) and other genes potentially involved in adaptation to the environment (e.g. genes related to iron acquisition and to antifungal resistance) were searched in each species. Gene cluster associated with secondary metabolites were searched with antiSMASH v.4 fungal version (FungiSMASH) [116] and SMURF [117].

### Antifungal susceptibility

*M. lusitanicus*, *M. endophyticus*, *M. fuscus*, *M. lanceolatus* and *M. racemosus*, as well as *Byssochlamys fulva* UBOCC-A-101005 (as a positive control), were grown in 5.4 cm Petri dishes on solid RPMI medium (Gibco BRL, Gaithersburg, MD) with 2% glucose and different concentrations of azole antifungal compounds: 0 $mg.l^{-1}$, 10 $mg.l^{-1}$, 50 $mg.l^{-1}$, 100 $mg.l^{-}$ 1, 500 $mg.l^{-1}$ and 1000 $mg.l^{-1}$ of voriconazole (VCZ) (OHRE Pharam, Tours, France) or posaconazole (PCZ) (MSD France, Courbevoie, France), respectively. Each antifungal concentration was assayed in triplicates. Growth diameters were measured after 7 days.

### Availability of supporting data

Raw sequence data were deposited at the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena/data/view/PRJEB30975) and genome assemblies, annotations and genome browser are available on the ABiMS platform (http://application.sb-roscoff.fr/project/mucor/) as well as on the JGI Mycocosm platform (https://genome.jgi.doe.gov/programs/fungi/index.jsf). Noteworthy, functional annotations found on the Mycocosm platform differ from the ones exploited here since the structural annotation was functionally re-annotated by the pipelines associated with the Mycocosm platform (https://genome.jgi.doe.gov/Mucrac1, https://genome.jgi.doe.gov/Mucend1, https://genome.jgi.doe.gov/Mucfus1, https://genome.jgi.doe.gov/Muclan1). All the genome sequences compared to the genomes sequenced in the present study were found in NCBI Genome except for *M. lusitanicus* genome available at the Mycocosm platform.

## Supplementary information

**Supplementary information** accompanies this paper at https://doi.org/10.1186/s12864-019-6256-2.

**Additional file 1: Figure S1.** Phylogenomic tree for the genome of 25 *Mucoromycotina*. The tree was reconstructed using Clann based on 29 genes trees. Each tree corresponded to one of the single copy gene families included at least 20 of the 25 isolates investigated. Bootstrap supports are indicated under the branch.

**Additional file 2: Figure S2.** Representation of the TE coverage in each genome depending on the TE category (classification of Wicker et al.) [118]. The four species sequenced in this study are in red. DHX: Helitron transposon. DTX: TIR transposon. DXX-MITE: unknown non-autonomous transposon, MITE-like. noCat: potential transposable element that could not be identified. RIX: LINE retrotransposons. RLX: LTR retrotransposon. RXX unknown retrotransposon. RXX-LARD: unknown non-autonomous retrotransposon, LARD-like. RXX-TRIM: unknown non-autonomous retrotransposon, TRIM-like.

**Additional file 3: Figure S3.** Number of CAZymes in the different fungal phyla. Major CAZymes classes are shown separately, Auxiliary redox enzymes (AA), Carbohydrate-Binding Modules (CBM), Carbohydrate Esterases (CE), Glycoside Hydrolases (GH), Glycoside Transferases (GT) and Polysaccharide Lyases (PL).

**Additional file 4: Table S1.** Number of identified CAZyme encoding genes involved in the degradation of plant cell wall components (Cellulose active, Hemicellulose active and Pectin active). *The presence of genes has been confirmed with manual annotation.*

**Additional file 5: Table S2.** Extraction and sequencing information corresponding to the four newly sequenced *Mucor* isolates

### Abbreviations

CAZymes: Carbohydrates Active enZymes; CYP51: 14α-demethylase paralogues; DMATS: DiMethylAllyl Tryptophan Synthase; FAS: Fatty Acid Synthase; NRPS: NonRibosomal Peptide Synthetase; PCZ: Posaconazole; PKS: Polyketide Synthase; RIA: Reductive Iron Assimilation; SSP: Small secreted protein; TPS: Terpene Synthase; VCZ: Voriconazole

### Acknowledgements

The authors are thankful to Vincent Bruno, for providing the annotation of the *Mucor* CDC-B8987, CDC-B7402, CDC-B9645, CDC-B5328 and CDC-B9738 isolates, to Jonathan Dorival, for advices on PKS, Antoine Branca and Emmanuelle Morin for stimulating discussion, Antoine Hermet for his involvement at the early stage of the project and Stephen Mondo for the integration of *Mucor* genomes in the Mycocosm platform. The authors are also thankful to the two anonymous reviewers for their constructive comments. The authors thank the UBOCC strain collection (https://www.univ-brest.fr/ubocc) for providing *Mucor* isolates. This research was funded by the Région Bretagne (ARED program) and EQUASA, a technological platform of the Université de Bretagne Occidentale, in the framework of the MUCORSCOPE project. The M. *fuscus, M. lanceolatus and M. racemosus* genome sequences were obtained in the context of the French National Agency for Research (ANR)project "Food-Microbiomes".

### Authors' contributions

Research idea conception: ECn, GB, LMC and JLJ. DNA Extraction: AL and RD. Antifungal assay: CLM. Analytical approach design: ECe, JLJ and LMC. Apollo instance deployment for manual annotation: LG and MM. Manual annotation: AL, JLJ, CPA, VG and LMC. Data analysis: AL, ECe, JLJ and LMC. Manuscript writing: AL, JLJ and LMC. Manuscript discussion and revision: ECn, GB and ECe. All authors read and approved the final manuscript.

### Funding

This research was funded by the the Région Bretagne (ARED program) and EQUASA, a technological platform of the Université de Bretagne Occidentale, in the framework of the MUCORSCOPE project.

### Availability of data and materials

Raw sequence data were deposited at the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena/data/view/PRJEB30975), genome assemblies and annotations are available on the ABiMS platform (http://application.sb-roscoff.fr/project/mucor/).

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

[1]Univ Brest, Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, F-29280 Plouzané, France. [2]Station Biologique de Roscoff, Plateforme ABiMS, CNRS: FR2424, Sorbonne Université (UPMC), Paris VI, Place Georges Teissier, 74 29682 Roscoff Cedex, BP, France. [3]CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), Sorbonne Université, 29680 Roscoff, France. [4]Department of Genetics and Microbiology, Faculty of Biology, University of Murcia, 30100 Murcia, Spain. [5]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, University Paris-Sud, Université Paris-Saclay, CEDEX 91198 Gif-sur-Yvette, France.

Received: 13 May 2019 Accepted: 31 October 2019
Published online: 10 February 2020

### References

1. Spatafora JW, Aime MC, Grigoriev IV, Martin F, Stajich JE, Blackwell M. The fungal tree of life: From molecular systematics to genomescale phylogenies. Microbiology Spectrum. 2017;5(5). https://doi.org/10.1128/microbiolspec.FUNK-0053-2016 Retrieved from https://escholarship.org/uc/item/4485m01m.
2. Morin-Sardin S, Nodet P, Coton E, Jany J-L. Mucor: a Janus-faced fungal genus with human health impact and industrial applications. Fungal Biol Rev. 2017;31(1):12–32.
3. Voigt K, Wolf T, Ochsenreiter K, Nagy G, Kaerger K, Shelest E, Papp T. In: Hoffmeister D, editor. 15 Genetic and Metabolic Aspects of Primary and Secondary Metabolism of the Zygomycetes. Cham: Springer International Publishing; 2016. p. 361–85.
4. Wagner L, Stielow B, Hoog S, Schwartze V, Kurzai O, Walther G. A new species concept for the clinically relevant Mucor circinelloides complex. Persoonia - Molecular Phylogeny and Evolution of Fungi; 2019.
5. Petrikkos G, Skiada A, Lortholary O, Roilides E, Walsh TJ, Kontoyiannis DP. Epidemiology and Clinical Manifestations of Mucormycosis. Clin Infect Dis. 2012;54(suppl_1):S23–34.
6. Pitt JI, Hocking AD. Fungi and food spoilage. US: Springer; 2009.

7. Morin-Sardin S, Rigalma K, Coroller L, Jany JL, Coton E. Effect of temperature, pH, and water activity on Mucor spp. growth on synthetic medium, cheese analog and cheese. Food Microbiol. 2016;56:69–79.
8. Orlowski M. Mucor dimorphism. Microbiol Rev. 1991;55(2):234–58.
9. Ma L-J, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, Elias M, Idnurm A, Lang BF, Sone T, et al. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-genome duplication. PLoS Genet. 2009;5(7):e1000549.
10. Khan MAK, Yang J, Hussain SA, Zhang H, Liang L, Garre V, Song Y. Construction of DGLA producing cell factory by genetic modification of Mucor circinelloides. Microb Cell Factories. 2019;18(1):64.
11. Vongsangnak W, Kingkaw A, Yang J, Song Y, Laoteng K. Dissecting metabolic behavior of lipid over-producing strain of Mucor circinelloides through genome-scale metabolic network and multi-level data integration. Gene. 2018;670:87–97.
12. Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM, Grimwood J, Álvarez MI, Avalos J, Bauer D, et al. Expansion of signal transduction pathways in fungi by extensive genome duplication. Curr Biol. 2016;26(12):1577–84.
13. Chibucos MC, Soliman S, Gebremariam T, Lee H, Daugherty S, Orvis J, Shetty AC, Crabtree J, Hazen TH, Etienne KA, et al. An integrated genomic and transcriptomic survey of mucormycosis-causing fungi. Nat Commun. 2016;7:1–11.
14. Liu M, Lin L, Gebremariam T, Luo G, Skory CD, French SW, Chou T-F, Edwards JE Jr, Ibrahim AS. Fob1 and Fob2 proteins are virulence determinants of Rhizopus oryzae via facilitating Iron uptake from Ferrioxamine. PLoS Pathog. 2015;11(5):e1004842.
15. López-Fernández L, Sanchis M, Navarro-Rodríguez P, Nicolás FE, Silva-Franco F, Guarro J, Garre V, Navarro-Mendoza MI, Pérez-Arques C, Capilla J. Understanding Mucor circinelloides pathogenesis by comparative genomics and phenotypical studies. Virulence. 2018;9(1):707–20.
16. López-Muñoz A, Nicolás FE, García-Moreno D, Pérez-Oliva AB, Navarro-Mendoza MI, Hernández-Oñate MA, Herrera-Estrella A, Torres-Martínez S, Ruiz-Vázquez RM, Garre V, et al. An adult Zebrafish model reveals that Mucormycosis induces apoptosis of infected macrophages. Sci Rep. 2018;8(1):12802.
17. Navarro-Mendoza MI, Pérez-Arques C, Murcia L, Martínez-García P, Lax C, Sanchis M, Capilla J, Nicolás FE, Garre V. Components of a new gene family of ferroxidases involved in virulence are functionally specialized in fungal dimorphism. Sci Rep. 2018;8(1):7660.
18. Patino-Medina JA, Maldonado-Herrera G, Perez-Arques C, Alejandre-Castaneda V, Reyes-Mares NY, Valle-Maldonado MI, Campos-Garcia J, Ortiz-Alvarado R, Jacome-Galarza IE, Ramirez-Diaz MI, et al. Control of morphology and virulence by ADP-ribosylation factors (Arf) in Mucor circinelloides. Curr Genet. 2018;64(4):853–69.
19. Patiño-Medina JA, Valle-Maldonado MI, Maldonado-Herrera G, Pérez-Arques C, Jácome-Galarza IE, Díaz-Pérez C, Díaz-Pérez AL, Araiza-Cervantes CA, Villagomez-Castro JC, Campos-García J, et al. Role of Arf-like proteins (Arl1 and Arl2) of Mucor circinelloides in virulence and antifungal susceptibility. Fungal Genet Biol. 2019;129:40–51.
20. Pérez-Arques C, Navarro-Mendoza MI, Murcia L, Lax C, Martínez-García P, Heitman J, Nicolás FE, Garre V. Mucor circinelloides Thrives inside the Phagosome through an Atf-Mediated Germination Pathway. mBio. 2019; 10(1):e02765–18.
21. Trieu TA, Navarro-Mendoza MI, Pérez-Arques C, Sanchis M, Capilla J, Navarro-Rodriguez P, Lopez-Fernandez L, Torres-Martínez S, Garre V, Ruiz-Vázquez RM, et al. RNAi-based functional genomics identifies new virulence determinants in Mucormycosis. PLoS Pathog. 2017;13(1):e1006150.
22. Álvarez E, Cano J, Stchigel AM, Sutton DA, Fothergill AW, Salas V, Rinaldi MG, Guarro J. Two new species of Mucor from clinical samples. Med Mycol. 2011;49(1):62–72.
23. Lebreton A, Meslet-Cladière L, Morin-Sardin S, Coton E, Jany J-L, Barbier G, Corre E. Comparative analysis of five Mucor species transcriptomes. Genomics. 2018;S0888-7543(18):30204.
24. Walther G, Pawłowska J, Alastruey-Izquierdo A, Wrzosek M, Rodriguez-Tudela JL, Dolatabadi S, Chakrabarti A, de Hoog GS. DNA barcoding in Mucorales: an inventory of biodiversity. Persoonia. 2013;30(1):11–47.
25. Gryganskyi AP, Golan J, Dolatabadi S, Mondo S, Robb S, Idnurm A, Muszewska A, Steczkiewicz K, Masonjones S, Liao H-L, et al. Phylogenetic and Phylogenomic Definition of Rhizopus Species. G3. 2018;8(6):2007–18.
26. Hermet A, Meheust D, Mounier J, Barbier G, Jany JL. Molecular systematics in the genus Mucor with special regards to species encountered in cheese. Fungal Biol. 2012;116(6):692–705.
27. Zheng R, Jiang H. Rhizomucor endophyticus sp.nov., an endophytic zygomycetes from higher plants. Mycotaxon. 1995;56:455–66.
28. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Program NCS, et al. Human skin fungal diversity. Nature. 2013;498(7454):367–70.
29. Lee SC, Billmyre RB, Li A, Carson S, Sykes SM, Huh EY, Mieczkowski P, Ko DC, Cuomo CA, Heitman J. Analysis of a Food-Borne Fungal Pathogen Outbreak: Virulence and Genome of a Mucor circinelloides Isolate from Yogurt. mBio. 2014;5(4):e01390–14.
30. Ibrahim AS, Spellberg B, Walsh TJ, Kontoyiannis DP. Pathogenesis of Mucormycosis. Clin Infect Dis. 2012;54(suppl_1):S16–22.
31. Thieken A, Winkelmann G. Rhizoferrin: a complexone type siderophore of the Mucorales and entomophthorales (Zygomycetes). FEMS Microbiol Lett. 1992;73(1–2):37–41.
32. Franken ACW, Lechner BE, Werner ER, Haas H, Lokman BC, Ram AFJ, van den Hondel CAMJJ, de Weert S, Punt PJ. Genome mining and functional genomics for siderophore production in Aspergillus Niger. Brief Funct Genomics. 2014;13(6):482–92.
33. Carroll CS, Grieve CL, Murugathasan I, Bennet AJ, Czekster CM, Liu H, Naismith J, Moore MM. The rhizoferrin biosynthetic gene in the fungal pathogen Rhizopus delemar is a novel member of the NIS gene family. Int J Biochem Cell Biol. 2017;89:136–46.
34. Ramakrishnan G. Iron and virulence in Francisella tularensis. Front Cell Infect Microbiol. 2017;7:107.
35. Caramalho R, Tyndall JDA, Monk BC, Larentis T, Lass-Flörl C, Lackner M. Intrinsic short-tailed azole resistance in mucormycetes is due to an evolutionary conserved aminoacid substitution of the lanosterol 14α-demethylase. Sci Rep. 2017;7(1):–15898.
36. Tang X, Chen H, Chen YQ, Chen W, Garre V, Song Y, Ratledge C. Comparison of biochemical activities between high and low lipid-producing strains of Mucor circinelloides: an explanation for the high Oleaginicity of strain WJ11. PLoS One. 2015;10(6):e0128396.
37. Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Pérez G, Pisabarro AG, Grigoriev IV, et al. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. PLoS Genet. 2016;12(6):e1006108.
38. Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2013;14(1):274.
39. Heitman J, Howlett BJ, Crous PW, Stukenbrock EH, James TY, Gow NAR. The fungal kingdom: American Society of Microbiology; 2017.
40. Pain A, Hertz-Fowler C. Genomic adaptation: a fungal perspective. Nat Rev Microbiol. 2008;6:572.
41. Plett JM, Martin F. Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. Trends Genet. 2011;27(1):14–22.
42. Knapp DG, Németh JB, Barry K, Hainaut M, Henrissat B, Johnson J, Kuo A, Lim JHP, Lipzen A, Nolan M, et al. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. Sci Rep. 2018;8(1):6321.
43. Gazis R, Kuo A, Riley R, LaButti K, Lipzen A, Lin J, Amirebrahimi M, Hesse CN, Spatafora JW, Henrissat B, et al. The genome of Xylona heveae provides a window into fungal endophytism. Fungal Biol. 2016;120(1):26–42.
44. Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet. 2015;47:410.
45. Karimi K, Zamani A. Mucor indicus: biology and industrial application perspectives: a review. Biotechnol Adv. 2013;31(4):466–81.
46. Kogan TV, Jadoun J, Mittelman L, Hirschberg K, Osherov N. Involvement of secreted Aspergillus fumigatus proteases in disruption of the actin Fiber cytoskeleton and loss of focal adhesion sites in infected A549 lung Pneumocytes. J Infect Dis. 2004;189(11):1965–73.
47. Shankar J, Tiwari S, Shishodia SK, Gangwar M, Hoda S, Thakur R, Vijayaraghavan P. Molecular insights into development and virulence determinants of Aspergilli: a proteomic perspective. Front Cell Infect Microbiol. 2018;8:180.
48. Rokas A, Wisecaver JH, Lind AL. The birth, evolution and death of metabolic gene clusters in fungi. Nat Rev Microbiol. 2018;16:731–44.
49. Bairwa G, Hee Jung W, Kronstad JW. Iron acquisition in fungal pathogens of humans. Metallomics. 2017;9(3):215–27.
50. Haas H. Iron - a key Nexus in the virulence of Aspergillus fumigatus. Front Microbiol. 2012;3:28.

51. Monnet C, Loux V, Gibrat J-F, Spinnler E, Barbe V, Vacherie B, Gavory F, Gourbeyre E, Siguier P, Chandler MJPO. The Arthrobacter arilaitensis Re117 genome sequence reveals its genetic adaptation to the surface of cheese. PLoS One. 2010;5(11):e15489.
52. Andrews JH. Fungal life-history strategies. Fungal Commun. 1992;2:119–45.
53. Cooke RC, Rayner AD. Ecology of saprotrophic fungi: Longman; 1984.
54. Howard DH. Pathogenic fungi in humans and animals. New York: Marcel Dekker; 2003.
55. Herbst DA, Townsend CA, Maier T. The architectures of iterative type I PKS and FAS. Nat Prod Rep. 2018;35:1046–69.
56. Maier T, Leibundgut M, Boehringer D, Ban N. Structure and function of eukaryotic fatty acid synthases. Q Rev Biophys. 2010;43(3):373–422.
57. Hitchman TS, Schmidt EW, Trail F, Rarick MD, Linz JE, Townsend CA. Hexanoate synthase, a specialized type I fatty acid synthase in Aflatoxin B1 biosynthesis. Bioorg Chem. 2001;29(5):293–307.
58. Csernetics Á, Nagy G, Iturriaga EA, Szekeres A, Eslava AP, Vágvölgyi C, Papp T. Biology: Expression of three isoprenoid biosynthesis genes and their effects on the carotenoid production of the zygomycete Mucor circinelloides. Fungal Genet Biol. 2011;48(7):696–703.
59. Navarro E, Sandmann G, Torres-Martínez S. Mutants of the carotenoid biosynthetic pathway of Mucor circinelloides. Exp Mycol. 1995;19(3):186–90.
60. Zhang Y, Navarro E, Cánovas-Márquez JT, Almagro L, Chen H, Chen YQ, Zhang H, Torres-Martínez S, Chen W, Garre V. A new regulatory mechanism controlling carotenogenesis in the fungus Mucor circinelloides as a target to generate β-carotene over-producing strains by genetic engineering. Microb Cell Factories. 2016;15:99.
61. Ropars J, Lo Y-C, Dumas E, Snirc A, Begerow D, Rollnik T, Lacoste S, Dupont J, Giraud T, López-Villavicencio M. Fertility depression among cheese-making Penicillium roqueforti strains suggests degeneration during domestication. Evolution. 2016;70(9):2099–109.
62. Winkelmann G. Specificity of iron transport in bacteria and fungi. In: Handbook of Microbial Iron Chelates. Boca Raton: CRC Press; 1991. p. 73–114.
63. Khan A. Synthesis, nature and utility of universal iron chelator – siderophore: a review. Microbiol Res. 2017;212:103–11.
64. Szebesczyk A, Olshvang E, Shanzer A, Carver PL, Gumienna-Kontecka E. Harnessing the power of fungal siderophores for the imaging and treatment of human diseases. Coord Chem Rev. 2016;327-328: 84–109.
65. Graninger W, Diab-Elschahawi M, Presterl E. Antifungal Agents. In: Presterl E, editor. Clinically Relevant Mycoses: A Practical Approach. Cham: Springer International Publishing; 2019. p. 31–42.
66. Woolley DW. Some new aspects of the relationship of chemical structure to biological activity. Science. 1944;100(2609):579–83.
67. Ozturkoglu-Budak S, Wiebenga A, Bron PA, de Vries RP. Protease and lipase activities of fungal and bacterial strains derived from an artisanal raw ewe's milk cheese. Int J Food Microbiol. 2016;237:17–27.
68. Falkiewicz-Dulík M. 6.8 - Leather and leather products. In: Falkiewicz-Dulik M, Janda K, Wypych G, editors. Handbook of Material Biodegradation, Biodeterioration, and Biostablization. 2nd ed. Toronto: ChemTec Publishing; 2015. p. 133–256.
69. Joichi Y, Chijimatsu I, Yarita K, Kamei K, Miki M, Onodera M, Harada M, Yokozaki M, Kobayashi M, Ohge H. Detection of Mucor velutinosus in a blood culture after autologous peripheral blood stem cell transplantation : a pediatric case report. Med Mycol J. 2014;55(2):E43–8.
70. Singh P, Paul S, Shivaprakash MR, Chakrabarti A, Ghosh AK. Stress response in medically important Mucorales. Mycoses. 2016;59(10):628–35.
71. Taj-Aldeen SJ, Almaslamani M, Theelen B, Boekhout T. Phylogenetic analysis reveals two genotypes of the emerging fungus Mucor indicus, an opportunistic human pathogen in immunocompromised patients. Emerg Microbes Infect. 2017;6(7):e63.
72. Fulton TM, Chunwongse J, Tanksley SD. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report. 1995;13(3):207–9.
73. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E, Debuchy R, et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun. 2014;5:2876.
74. Andrews S. FASTQC: A quality control tool for high throughput sequence data; 2010.
75. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10–2.
76. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
77. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.
78. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 2012;1:18.
79. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and Phylogenomics. Mol Biol Evol. 2018; 35(3):543–8.
80. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008;18(12):1979–90.
81. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7(1):62.
82. Dobin A, Gingeras TR. Mapping RNA-seq reads with STAR. Curr Protoc Bioinformatics. 2015;51:11.14.11–9.
83. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nat Biotechnol. 2010;28(5):511–5.
84. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011; 29(7):644–52.
85. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21(9):1859–75.
86. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6(1):31.
87. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to assemble spliced alignments. Genome Biol. 2008;9(1):R7.
88. Krogh A, Larsson B, von Heijne G, Sonnhammer E. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80.
89. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785.
90. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.
91. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):D279–85.
92. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
93. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 2014;42(D1):D699–704.
94. Claudel-Renard C, Chevalet C, Faraut T, Kahn D. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res. 2003;31(22):6633–9.
95. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44(Web Server issue):W54–7.
96. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.
97. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933–5.
98. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, et al. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 2015;43(Database issue):D130–7.
99. Dunn N A, Unni D, Diesh C, Munoz-Torres M, Harris N L, Yao E, Rasche H, Holmes I H, Elsik C G, Lewis S E: Apollo: democratizing genome annotation. PLoS Comput Biol. 2019;15(2):e1006790.
100. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in De novo annotation approaches. PLoS One. 2011;6(1):e16526.
101. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, Quesneville H. PASTEC: an automatic transposable element classification tool. PLoS One. 2014;9(5):e91929.

102. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95–W101.
103. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490–5.
104. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2014;12:59.
105. Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. BMC Bioinformatics. 2017;18(1):214.
106. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2016; 44(D1):D343–50.
107. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157.
108. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155–70.
109. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
110. Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 2013;62(4):611–5.
111. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
112. Creevey CJ, McInerney JO. Clann: investigating phylogenetic information through supertree analyses. Bioinformatics. 2004;21(3):390–2.
113. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 2003; 19(2):301–2.
114. Zhou P, Zhang G, Chen S, Jiang Z, Tang Y, Henrissat B, Yan Q, Yang S, Chen C-F, Zhang B, et al. Genome sequence and transcriptome analyses of the thermophilic zygomycete fungus Rhizomucor miehei. BMC Genomics. 2014; 15(1):294.
115. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 2013;30(8):1987–97.
116. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. 2017;45(W1):W36–41.
117. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 2010;47(9):736–41.
118. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8(12):973–82.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

**Learn more** biomedcentral.com/submissions

BMC