



Gene Expansion Shapes Genome Architecture in the Human Pathogen *Lichtheimia corymbifera*: An Evolutionary Genomics Analysis in the Ancient Terrestrial Mucorales (Mucoromycotina)

Volker U. Schwartze^{1,2,3}, Sascha Winter^{3,9}, Ekaterina Shelest^{4,9}, Marina Marcet-Houben^{5,6,9}, Fabian Horn^{4,9}, Stefanie Wehner³, Jörg Linde⁴, Vito Valiante², Michael Sammeth^{7,8}, Konstantin Riege³, Minou Nowrousian⁹, Kerstin Kaerger^{1,2}, Ilse D. Jacobsen¹⁰, Manja Marz³, Axel A. Brakhage^{1,2,11}, Toni Gabaldón^{5,6,11}, Sebastian Böcker^{3,11}, Kerstin Voigt^{1,2,11}*

1 University of Jena, Institute of Microbiology, Department of Microbiology and Molecular Biology, Jena, Germany, **2** Leibniz Institute for Natural Product Research and Infection Biology, Department of Molecular and Applied Microbiology, Hans Knöll Institute, Jena, Germany, **3** University of Jena, Department of Bioinformatics, Jena, Germany, **4** Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Systems Biology/Bioinformatics, Jena, Germany, **5** Centre for Genomic Regulation (CRG), Barcelona, Spain, **6** Universitat Pompeu Fabra (UPF), Barcelona, Spain, **7** Centre Nacional d'Anàlisi Genòmica (CNAG), Functional Bioinformatics, Barcelona, Spain, **8** Laboratório Nacional de Computação Científica (LNCC), Petrópolis, Rio de Janeiro, Brazil, **9** Ruhr University Bochum, Department of General and Molecular Botany, Bochum, Germany, **10** Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Department of Microbial Immunology, Jena, Germany

Abstract

Lichtheimia species are the second most important cause of mucormycosis in Europe. To provide broader insights into the molecular basis of the pathogenicity-associated traits of the basal Mucorales, we report the full genome sequence of *L. corymbifera* and compared it to the genome of *Rhizopus oryzae*, the most common cause of mucormycosis worldwide. The genome assembly encompasses 33.6 MB and 12,379 protein-coding genes. This study reveals four major differences of the *L. corymbifera* genome to *R. oryzae*: (i) the presence of an highly elevated number of gene duplications which are unlike *R. oryzae* not due to whole genome duplication (WGD), (ii) despite the relatively high incidence of introns, alternative splicing (AS) is not frequently observed for the generation of paralogs and in response to stress, (iii) the content of repetitive elements is strikingly low (<5%), (iv) *L. corymbifera* is typically haploid. Novel virulence factors were identified which may be involved in the regulation of the adaptation to iron-limitation, e.g. LCor01340.1 encoding a putative siderophore transporter and LCor00410.1 involved in the siderophore metabolism. Genes encoding the transcription factors LCor08192.1 and LCor01236.1, which are similar to GATA type regulators and to calcineurin regulated CRZ1, respectively, indicating an involvement of the calcineurin pathway in the adaption to iron limitation. Genes encoding MADS-box transcription factors are elevated up to 11 copies compared to the 1–4 copies usually found in other fungi. More findings are: (i) lower content of tRNAs, but unique codons in *L. corymbifera*, (ii) Over 25% of the proteins are apparently specific for *L. corymbifera*. (iii) *L. corymbifera* contains only 2/3 of the proteases (known to be essential virulence factors) in comparison to *R. oryzae*. On the other hand, the number of secreted proteases, however, is roughly twice as high as in *R. oryzae*.

Citation: Schwartze VU, Winter S, Shelest E, Marcet-Houben M, Horn F, et al. (2014) Gene Expansion Shapes Genome Architecture in the Human Pathogen *Lichtheimia corymbifera*: An Evolutionary Genomics Analysis in the Ancient Terrestrial Mucorales (Mucoromycotina). PLoS Genet 10(8): e1004496. doi:10.1371/journal.pgen.1004496

Editor: Ashraf S. Ibrahim, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, United States of America

Received: January 2, 2013; **Accepted:** May 24, 2014; **Published:** August 14, 2014

Copyright: © 2014 Schwartze et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: TG group is funded in part by a grant from the Spanish Ministry of Science and Innovation (BFU2009-09168). MM group was in part supported by the German Science Foundation (DFG) MA5082/1-1 (KR). MN would like to thank Ulrich Kück for generous support, and the DFG for funding (project NO407/4-1). SB group was in part supported by DFG BO 1910/8-1 (SWi). KV and AAB acknowledge financial support of the Leibniz Institute of Natural Product Research and Infection Biology - Hans Knöll Institute Jena, Germany. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: Kerstin.Voigt@hki-jena.de

These authors have contributed equally to this work.

These authors share senior authorship on this work.

Introduction

The basal lineages of terrestrial fungi, formerly Zygomycota, were recently shown to be polyphyletic and were therefore separated into four separate subphyla [1]. Especially the order

Mucorales of the Mucoromycotina encompasses several human pathogenic species. Although infections with mucoralean fungi (mucormycosis) are less common as compared to aspergilloses or candidioses, these fungi are increasingly recognized as the source of infection in immunocompromised patients [2]. Mucormycoses

Author Summary

Lichtheimia species are ubiquitous saprophytic fungi, which cause life-threatening infections in humans. In contrast to the mucoralean pathogen *R. oryzae*, *Lichtheimia* species belong to the ancient mucoralean lineages. We determined the genome of *L. corymbifera* (formerly *Mycocladius corymbifer* ex *Absidia corymbifera*) and found high dissimilarities between *L. corymbifera* and other sequenced mucoralean fungi in terms of gene families and synteny. A highly elevated number of gene duplications and expansions was observed, which comprises virulence-associated genes like proteases, transporters and iron uptake genes but also transcription factors and genes involved in signal transduction. In contrast to *R. oryzae*, we did not find evidence for a recent whole genome duplication in *Lichtheimia*. However, gene duplications create functionally diverse paralogs in *L. corymbifera*, which are differentially expressed in virulence-related compared to standard conditions. In addition, new potential virulence factors could be identified which may play a role in the regulation of the adaptation to iron-limitation. The *L. corymbifera* genome and the phylome will advance further research and better understanding of virulence mechanisms of these medically important pathogens at the level of genome architecture and evolution.

are associated with rapid blood vessel invasion and massive destruction of tissue (necrosis) [3,4]. Mortality rates are high (~50%) and treatment mainly includes a combination of antifungals and extensive surgery [2,5–7]. In addition, mucoralean pathogens are resistant to a variety of antifungals including voriconazole which makes treatment even more complicated [8].

The order Mucorales comprises 240 described species, of which at least 20 have been found to be involved in mucormycosis. Genome sequences have been published for only two important pathogenic species within the Mucorales, namely *Rhizopus oryzae* (= *R. arrhizus*) and *Mucor circinelloides*. These species are closely related and represent derived lineages within the group. However, a large proportion of pathogenic Mucorales (10 species) belong to more basal groups including the genera *Lichtheimia*, *Rhizomucor*, *Apophysomyces*, *Saksenaea* and *Syncephalastrum*. Recently, the first report of the involvement of *Thamnostylum lucknowense*, an ancient mucoralean fungus, in human infections has been published [9]. To date, almost nothing is known about the genomic structure and pathogenicity mechanisms of these basal groups.

Lichtheimia species are ubiquitous saprophytic molds and represent the second and third most common cause of mucormycosis in Europe and worldwide, respectively [2,7,10,11]. The genus *Lichtheimia* was formerly included in the genus *Absidia* based on morphological similarities [12]. However, based on the higher growth optimum as well as morphological and molecular data *Lichtheimia* species were separated from the mesophilic *Absidia* species [13]. Today the genus encompasses five thermotolerant species, of which three are known to be clinically relevant, namely *L. corymbifera*, *L. ramosa* and *L. ornata* [14]. In addition to the distinct phylogenetic position at the base of mucoralean fungi, *Lichtheimia* species exhibit differences in physiology compared to the sequenced pathogens *M. circinelloides* and *R. oryzae*, including a higher maximum growth temperature (48–52°C vs 37°C and <45°C) and differences in susceptibility to certain antifungals [8,15]. Moreover, filamentously growing *Mucor* and *Rhizopus* species have been shown to be able to form

yeast cells which were also found in patient material and thus might be of relevance during infection [16–18]. In contrast, no yeast-like growth forms of *Lichtheimia* species have been observed to date. In addition, pulmonary *Lichtheimia* infections following solid organ transplantation seem to be associated with a higher risk to develop disseminated disease [19]. Besides its role in human infections, *L. corymbifera* is also believed to be associated with Farmer's lung disease (FLD), a hypersensitivity disorder resulting from frequent contact of mouldy material in agriculture [20]. Nothing comparable has been described for other mucoralean species. In addition to their pathogenicity towards humans, several *Lichtheimia* species are known as contaminants of several food products (e.g. cocoa, peanuts, olive products) [21–23]. However, despite the known role of *Lichtheimia* species in infection and diseases, several *Lichtheimia* species play an important role in the fermentation of soy products in Asian cuisine [24]. The large evolutionary distance and notable differences in infection strategies between *Lichtheimia* and the two sequenced mucoralean pathogens indicate that they independently evolved their ability to infect humans by developing specific pathogenesis mechanisms. To gain insight into the genomic differences between these groups of pathogens, here we report the genome sequence of the type-strain of *L. corymbifera* (FSU 9682, CBS 429.75, ATCC 46771) which has been shown to be a typical strain in terms of virulence and physiology for this species [25] and compare it to published genomes of mucoralean fungi and other fungal phyla.

Results/Discussion

Genome assembly and structure

The genome of the type-strain of *L. corymbifera* (FSU 9682, CBS 429.75, ATCC 46771) was sequenced by a combination of 454 sequencing of a shotgun and 8 kb paired-end library in combination with Illumina sequencing of a paired-end read library (Materials and methods, Table S1). The final assembly comprises 209 scaffolds with a N50 scaffold size of 367,562 nt and a total length of 33.6 Mb (Table 1), which is comparable to the genome size of other zygomycetous fungi [26]. Mucoralean genomes are generally believed to contain large amounts of repetitive elements representing around 35% of the genome [27]. However, analysis of the *L. corymbifera* genome shows a much smaller content of repetitive elements, with only 4.7% of the assembly representing repetitive elements including DNA transposons, LTR and non-LTR retrotransposons (Table S2). This finding is consistent with the results of the k-mer analyses on the Illumina reads where only low amounts of potential repetitive regions were found. Of note, all previous estimates of repetitive elements in mucoraleans correspond to species with large genomes such as *R. oryzae* (46 Mb; 20% repetitive elements), *Absidia glauca* (52 Mb; 35%) and *P. blakesleeianus* (54 Mb; 35% repetitive elements) [27,28]. Interestingly, a *Lichtheimia*-specific gene expansion in the heterokaryon incompatibility genes was discovered (see section gene expansion) which are involved in the recognition of non-self DNA and may contribute to the low amount of repetitive elements. Another mechanism of protection against transposons and viruses is RNA interference resulting in sequence specific RNA degradation [29]. Several predicted proteins with functional domains associated with this mechanism were found including a dicer-like protein, one argonaute-2 protein and a translation initiation factor 2C homolog. However, the exact effects of these mechanisms on the amount of repetitive elements remain to be determined.

Heterozygosity was shown for several fungi including the basal lineage fungus *Batrachochytrium dendrobatidis* [30]. In order to

Table 1. Statistics of the *L. corymbifera* genome.

Assembly statistics	
Total scaffold length (Mb)	33.6
Scaffolds	209
N50 contig length (nt)	66,718
N50 scaffold length (nt)	367,562
G+C content	43.4%
Predicted protein-coding genes	
Predicted genes	12,379
Average coding sequence size (nt)	1,287
Average G+C content in coding sequence	46.2%
Total introns	48,663
Introns per gene (median)	4.8
Average intron length (nt)	258
Predicted non-coding RNA genes	
Predicted genes	213
Average G+C content in non-coding RNAs	49.2%
Total introns	3

doi:10.1371/journal.pgen.1004496.t001

test for potential heterozygous regions and estimate the genome size of *L. corymbifera*, k-mer analyses based on the Illumina reads were performed using an algorithm described previously [31] (Material and Methods). Analysis resulted in a relatively clear single peak with a slightly trailing left flank for all k-values (Figure S1). The distribution could be dissected into three components, each showing a normal distribution with similar variance, but different means and different proportions. The main component represents the potential homozygous part of the genome (94%), whereas two small components represent the potential heterozygous part of the genome (4%), and most likely some repeat regions that occur at relatively low frequency (2%). It has to be noted that the potential heterozygous part is rather small and could as well be explained e.g. by regions that are difficult to sequence and therefore have lower k-mer coverage. The lack of heterozygosity is in accordance with the general assumption that mucoralean fungi are haploid during vegetative growth. Based on the k-mer analysis for different k-mer lengths (41, 59, 69, and 79 nt) a total genome size of around 35 Mb was predicted which is close to our total scaffold length of 33.6 Mb (96% of k-mer predicted size).

Non-coding RNA prediction and annotation

We annotated 174 tRNAs in *L. corymbifera*. Although *R. oryzae* (239) comprises many more tRNAs, we found unique anticodons among the basal fungi in *Lichtheimia*: CCC (Gly), AAA (Phe) and GAT (Ile). In contrast, only *L. corymbifera* misses the anticodons CAC (Val), CCT (Arg) and TAT (Ile). Three GTA (Tyr) tRNAs were predicted with introns in *L. corymbifera*, while the number was higher in other mucoralean fungi (up to 10). No selenocysteine and possible suppressor tRNAs were predicted. We found the downstream half of 28S rRNA only, but no 18S rRNA in the current assembly. We expect at least two operons (18S – 5.8S – 28S rRNA) as found in *R. oryzae*. In addition to 5S rRNAs located close to the operons, we were able to identify several independent 5S rRNA copies (Table S3). Another housekeeping ncRNA, present in all kingdoms of life, is the ribozyme RNase P, which processes tRNAs by cleaving off nucleotides on the 3' end of tRNAs [32]. We detected this gene as expected in a single copy per

genome, but two identical copies are apparently present in the genome of *R. oryzae*, which may result from whole genome duplication in *R. oryzae* [28]. The pseudoknot in the centre of the molecule is accredited with the catalytic function and highly conserved in evolution [33]. However, the *L. corymbifera* candidate varies exceptionally in sequence, while the secondary structure is maintained. Whether the function of the molecule is affected has to be analyzed. The evolutionary related RNase MRP was invented at the origin of eukaryotes with dual function: (a) initiation of mitochondrial replication and (b) separation of 18S rRNA from 5.8S rRNA [34]. One copy per basal fungal genome was detected. The signal recognition particle containing a ncRNA part (SRP RNA) guides proteins to the endoplasmatic reticulum [35]. One copy was detected in *Lichtheimia*, whereas two copies were identified in the genome of *R. oryzae*. Surprisingly, the covariance model of mucoralean fungi, in agreement with *Rhizopus*, *Batrachochytrium* and *Monosiga* (RF00017) is much closer related to metazoans than to other known fungi SRP RNAs. We detected the RNA components of the major spliceosome and collected indications for a functional minor spliceosome. Except for U4 snRNA all five RNAs involved in U2-splicing were detected in *Lichtheimia*. U4 snRNA was not part of the assembly; however an U4-candidate was identified in the originally sequenced read data. Additionally, four of five RNAs involved in AT-AC-splicing were found. However, several special secondary structures were discovered, which may alter the functionality of the minor spliceosome: (i) The third stem of U12 snRNA is atrophied and the last stem is shorter than expected for all basal fungi. (ii) U4atac is not detected in *Lichtheimia*. The other basal fungi show one inconspicuous copy, which is assumed to be an assembly mistake. However, no similar homologous gene was detected in reads either. (iii) The second half of U6atac is highly divergent (Figure S2 and supplemental material: <http://www.rna.uni-jena.de/supplements/lichtheimia/index.html>). Eleven CD-box snoRNAs and 3 H/ACA snoRNAs were identified, which are mainly conserved in sequence and structure among basal fungi. For further details we refer to the supplemental material (www.rna.uni-jena.de/supplements/lichtheimia/index.html). Additionally, several ncRNA candidates could be proposed, which have to be functionally characterized in future experiments. A riboswitch, binding to thiamine pyrophosphate (TPP) was found in all basal fungi. For *Lichtheimia* a potential telomerase RNA is suggested, which is surprisingly closely related to the shortest known telomerase RNAs in ciliates (150 nt *Tetrahymena paravorax*). This is unexpected, since the longest telomerase is known from the fungus *Saccharomyces cerevisiae* (1,220 nt). Although the alignment of the usually extremely divergent telomerase RNA is very convincing in sequence and secondary structure (see supplemental material), no homologs in another basal fungus and no interacting ciliate protein homolog were found in our current assembly. U7 snRNA is known to interact with the downstream region of histone mRNA for inhibition of degradation. Four similar candidates for this short ncRNA were identified. In eukaryotes, polymerase III transcripts (e.g. U6 snRNA, RNase P, RNase MRP, SRP RNA, U6atac snRNA) usually display a typical promoter region: –10 nt TATA box, PSE element, Oct region [36]. Therefore, a search for conserved motifs was conducted in *Lichtheimia* promoter regions. However, we were not able to identify even one of these motifs. This highlights a possible modified polymerase III activity for basal fungi and has to be investigated in detail in further work. A phylogeny among basal fungi and *Schizosaccharomyces pombe* as outgroup based on ncRNAs (except 18S and 28S rRNA) was reconstructed, see Figure S2 (B and C). In accordance with protein and traditional

rRNA phylogeny *Lichtheimia* groups basal to *P. blakesleeanus* and the other two investigated fungi.

Protein-coding gene prediction and annotation

To aid prediction of protein-coding genes, RNA-seq analyses were performed for three different growth conditions in three biological replicates (see Material and Methods). The use of different conditions should ensure a higher number of expressed genes, thereby allowing evidence-based gene predictions for many gene models. On average, each replicate has a 70-fold genome coverage, which sums up to a 630-fold genome coverage (Table S4). Prediction of protein-coding genes was performed using AUGUSTUS [37], resulting in 12,379 predicted genes. Genes were functionally annotated by comparing to GenBank sequences using BLASTp (E-value $\leq 10^{-25}$), and by scanning for the presence of conserved domains using the InterProScan function of BLAST2GO [38]. BLAST hits were obtained for 7,917 genes, InterProScan results were found in 10,066 genes and at least one Gene Ontology (GO) term was assigned to 7,435 genes based on the union of BLAST and InterProScan results. The raw reads of the DNA- and RNA-seq experiments, the final genome assembly, the structural and functional gene prediction are available at <http://www.ebi.ac.uk/ena/data/view/PRJEB3978>. The genome data are also accessible *via* HKI Genome Resource (<http://www.genome-resource.de/>).

Comparison of protein-coding genes between *L. corymbifera* and other completely sequenced genomes

An exhaustive comparison of *L. corymbifera* genome with other 24 completely sequenced genomes including the major fungal groups (Chytridiomycota, Mucoromycotina, Asco- and Basidiomycota) was performed. This comparison included the reconstruction of *L. corymbifera* phylome, which encompasses the complete set of evolutionary histories of *L. corymbifera* genes (Material and Methods). It was carried out using the previously described PhylomeDB pipeline [39]. In brief, for each *L. corymbifera* protein-coding gene we searched for homologs, and multiple sequence alignments were built, and Maximum Likelihood analyses were performed to reconstruct a phylogenetic tree. The phylome is available through phylomeDB (<http://phylomedb.org>), with the phylome ID 245. The phylome was used to establish phylogeny-based orthology and paralogy relationships among genes in the species considered [40], and to detect gene expansions (see below). In addition, we used two complementary approaches, gene concatenation and super-tree [41], to reconstruct the species tree that represents the evolution of the 25 species considered. In the first approach, 58 genes were selected that were present in 21 out of 25 in single copy. Their corresponding alignments were then concatenated and a maximum-likelihood species tree was reconstructed (Material and Methods). In the second approach, 9,478 trees present in the phylome were used to build a super-tree using a gene tree parsimony approach, a method which finds the topology that minimizes the total number of duplications in the phylome [42]. Both resulting trees presented a similar topology, which placed *L. corymbifera* at the base of the other Mucorales species (Figure 1). The only difference found between the trees by the complementary approaches was the position of *Schizosaccharomyces pombe*, which appeared at the base of Ascomycota in the super-tree tree while in the concatenated tree it grouped with *S. cerevisiae*. To assess the level of overlap in genetic content between the different species an all-against-all comparison of the 25 genomes was performed. The results indicate between 50% and 75% of the proteins encoded in the other three Mucorales

species had homologs in *L. corymbifera* (Figure 1). Surprisingly, this percentage of shared gene content with *L. corymbifera* was similar to that of *Schizosaccharomyces pombe* (60.6%), which is higher than that found in the more closely related chytrid *B. dendrobatidis* (an average of 41.1%). Figure 1 also shows how these homologs are distributed in differently defined groups. Most interestingly, the fraction of species-specific proteins (grey bars in the figure) is particularly high in large genomes (e.g., over half of the largest genomes *Laccaria bicolor* and *Puccinia graminis*). Over 25% of the proteins apparently are specific for *L. corymbifera*.

Conserved gene regions in *L. corymbifera* and other mucoralean genomes

Since the *Lichtheimia* lineage separated early in mucoralean evolution we can expect that severe genomic re-arrangements have taken place during evolution, causing substantial differences between the genome structures of *L. corymbifera* and other Mucorales. Only 57.7% of the gene families present in *Lichtheimia* are also present in at least one of the other mucoralean genomes while only 36.7% were found in all four genomes representing 70.4% and 53.7% of the total *L. corymbifera* genes, respectively (Figure 2 A). Conserved regions, in terms of gene order, between mucoralean genomes were examined and evaluated with respect to the amount of conserved genes of these regions. A total of 230 regions with a minimum of 3 conserved genes of *L. corymbifera* were found that were present in at least one of the other genomes. These regions were interspersed over 41.1% of the scaffolds but covered only 7.6% of the *L. corymbifera* genome reflecting the high dissimilarity between the mucoralean genomes (Figure 2 B). Only 6 regions were shared with all species. The total number of shared clusters was found to be consistent with the phylogenetic distance between the species (Figure 2 B+C). Genes in the conserved regions are members of different gene families and contain a variety of functional domains.

Gene duplications in *L. corymbifera*

In accordance with former results [28] a higher number of gene families with two members were detected for *R. oryzae* but also for *L. corymbifera* as compared to other fungi ($6.99\% \pm 0.58\%$) (Figure 3 A). Whole genome duplication has been previously described for *R. oryzae* based on the presence of gene duplications and duplication of large genomic regions (segmental duplications) [28]. To investigate whether segmental duplications and thus a potential WGD also occur in *L. corymbifera*, the genomes were scanned for the presence of duplicated regions using GECKO2 [43,44]. Consistent with the former findings, our analysis showed a high number of segmental duplications in *R. oryzae* covering more than 10% of the genome [28] while fewer duplicated regions were found in *L. corymbifera* covering less than 4% of its genome (Figure 3 B). Thus, the gene duplications seem not to result from recent WGD as in *R. oryzae* but may result from an ancient genome duplication in mucoralean fungi as suggested by Marcet-Houben et al. [45] which is no longer detectable in the duplicated gene clusters. The possibility of ancient WGD in mucoralean genomes is currently investigated in more detail (Corrochano et al., pers. comm.). The genome of *L. corymbifera* also shows increased numbers of gene families with a higher number of genes indicating that gene duplication and the preservation of the gene copies seem to be a common process in mucoralean genomes and may be independent from WGD. This will be further addressed in the next section.

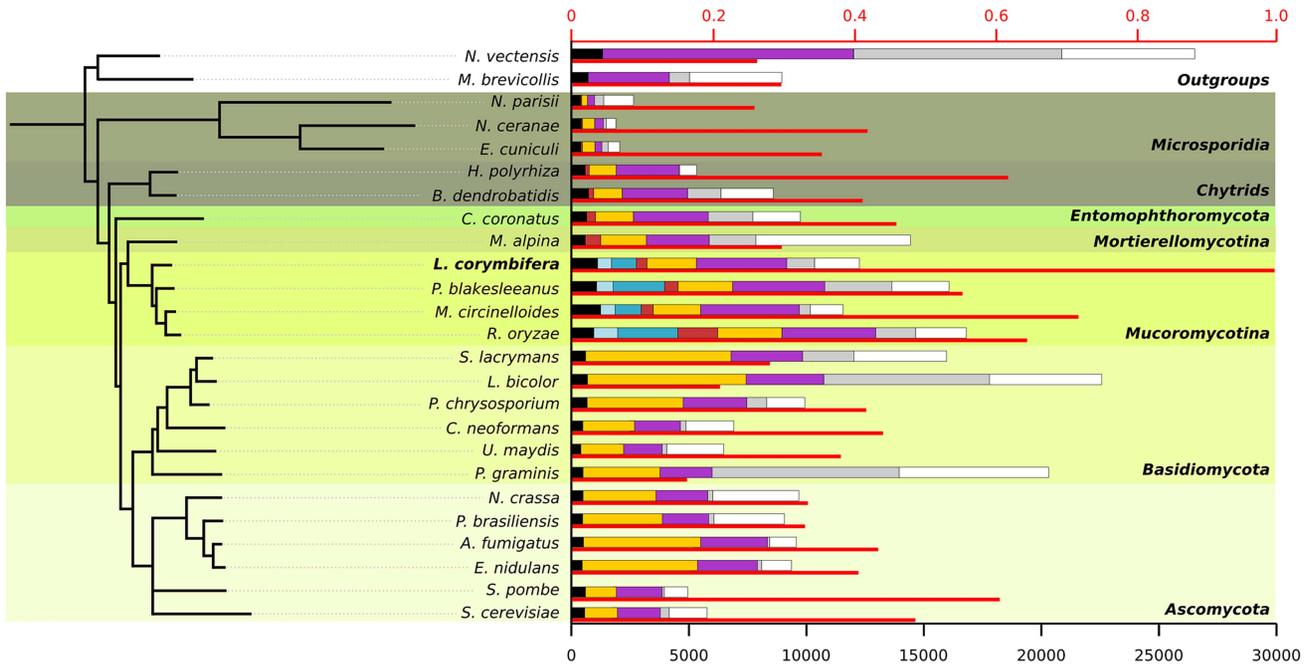


Figure 1. Species tree including the 25 species used during phylome reconstruction. For each species the thin red bar represents the proportion of proteins that have a homolog in *L. corymbifera* (upper axis). The coloured bars represent the number of proteins found in different ranges of species (lower axis): black: wide-spread proteins found in at least 23 of the 25 species, light blue: proteins found exclusively in all four Mucorales species, darker blue: proteins found only in Mucorales species, red: proteins found in early diverging fungi, yellow: proteins found in fungi and at least one of the outgroups, grey: species-specific proteins without orthologs in other species but with paralogs within the genome, white: proteins with no homologs. All nodes in the tree have a bootstrap support of 100.
doi:10.1371/journal.pgen.1004496.g001

Gene expansions in the *L. corymbifera* genome

In addition to gene duplications shared by all mucoralean fungi a high amount of species-specific duplications was detected. Therefore, the phylome was scanned in search of expansions of protein families that occurred specifically in *L. corymbifera*. For each tree, ETE [46] was used to find nodes that contained at least five *L. corymbifera* sequences and no other fungal sequence (Figure 4 A and B). Overlapping expansions were fused when they

shared more than half of their members. We found 75 expansions that fulfilled those requirements. Five expansions contained more than 30 members, with the largest containing 331 paralogous genes. In contrast, the large genome of *R. oryzae* contains approximately twice the number of expansions, with the largest encompassing 1,888 members. As some of those expansions are likely the result of the presence of transposons, we scanned them for the presence of transposon-linked domains using the Pfam

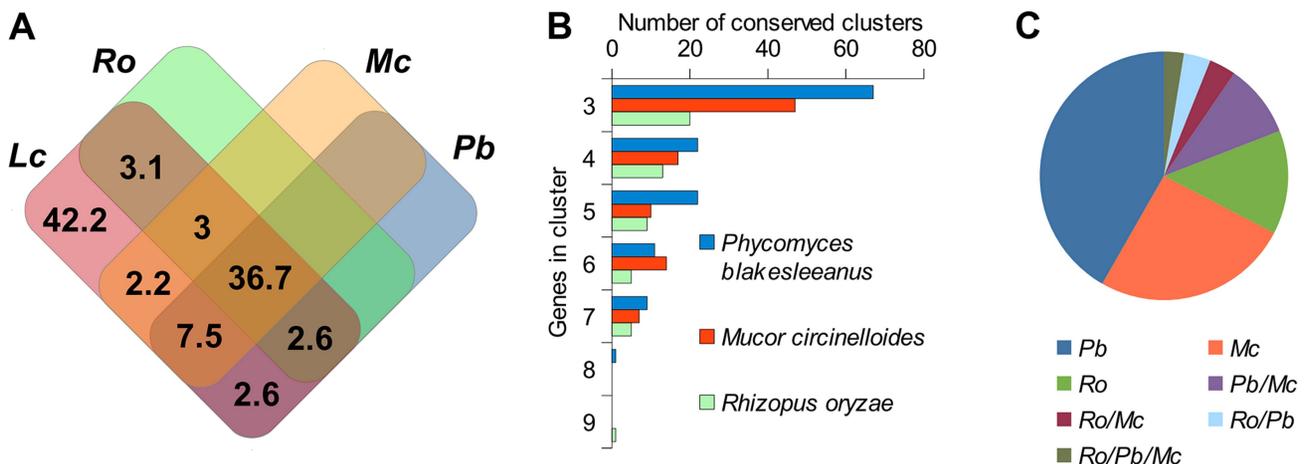


Figure 2. Conserved regions of the *L. corymbifera* genome with other mucoralean genomes. (A) Venn diagram of shared gene families between *L. corymbifera* and other mucoralean fungi based on GhostFam gene families. Numbers indicate percentage of *L. corymbifera* gene families. (B) Number and size of conserved clusters of *L. corymbifera* with other mucoralean genomes. (C) Proportions of conserved clusters of *L. corymbifera* shared with different mucoralean genomes. Occurrences in more than one of the genomes are indicated by a slash between the species.
doi:10.1371/journal.pgen.1004496.g002

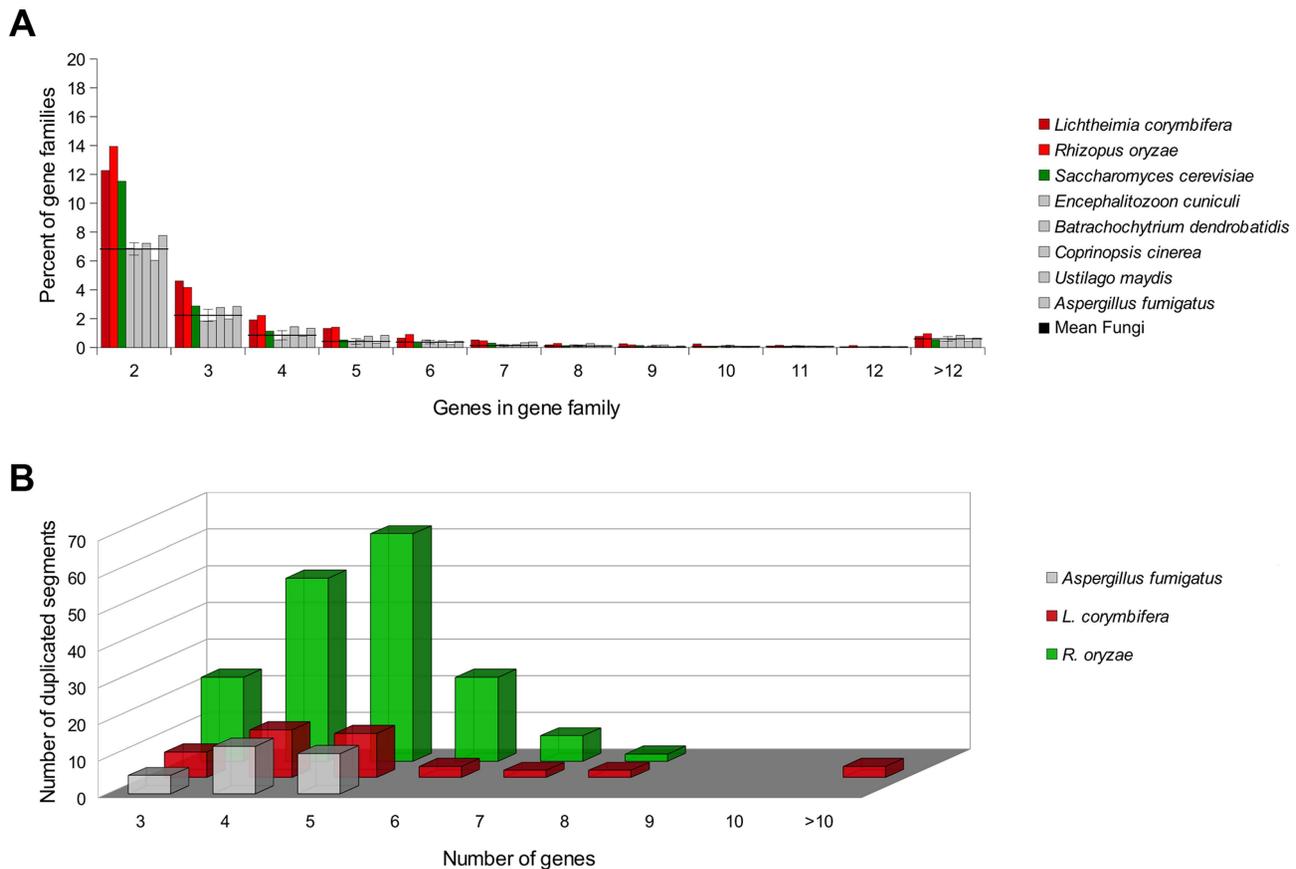


Figure 3. Gene duplication and duplication of genomic regions within mucoralean genomes in comparison to the genome of *A. fumigatus*, which (i) inhabits the same natural habitats and (ii) causes similar symptomatology in human like *L. corymbifera* and (iii) serves as model organism for causative agents of invasive mycoses [56,75]. The genome of *A. fumigatus* serves as measure for low incidences of singular and segmental gene duplications [115,116]. (A) Comparison of gene families between *L. corymbifera*, *R. oryzae* and non-mucoralean fungi. Gene families are based on GhostFam homology families. Values for *L. corymbifera* and *R. oryzae* are excluded from the "mean fungi" value. (B) Regions with a minimum of 3 genes were tested for multiple occurrences within the genomes by GECKO2. doi:10.1371/journal.pgen.1004496.g003

database and HMMER3 [47,48]. If the expansions that contain transposons were excluded, 66 groups of paralogous proteins will be left in *L. corymbifera* comprising a total of 820 genes (Figure 4 B).

The most abundant expanded groups (with 331 and 242 members) are rather heterogeneous in terms of functional domains, thus there is no particular function that could be assigned to them. The largest group with a dominating domain contains in total 56 members, of which 52 possess a heterokaryon incompatibility protein (HET) domain (PF06985) (Figure 4 C). Interestingly, this domain was so far attributed nearly exclusively to ascomycetes (with only one exception for the basidiomycete *Moniliophthora perniciosa*), where the HET proteins control somatic allorecognition (non-self-recognition) during the formation of heterokaryons [49]. However, Mucorales, opposite to ascomycetes, do not form heterokaryons by fusion of somatic cells but only during sexual reproduction and zygospore-formation. Since HET domain proteins are absent in all other sequenced zygomycetous genomes it is unlikely that they play a general role in the sexual reproduction but seem to be specific for *Lichtheimia*. Interestingly, these HET genes were differentially regulated under stress conditions. Several copies of the HET domain proteins were down-regulated under iron-depletion and hypoxia. Since these genes are absent in all other mucoralean fungi it is unclear which

functions they serve in *L. corymbifera*. In addition it is unclear where these genes originate since they do not occur in other basal fungi and show only very weak similarity with the HET proteins of dikaryan fungi.

Several expansions contain transporters: major facilitator superfamily (MFS, PF07690, PF12832, PF05977, PF13347), ABC transporters (PF00005, PF00501, PF01061, PF00664, PF06422), sugar (and other) transporters (PF00083). In addition, some interesting expansions are connected to the transcription regulation function, which is discussed in more detail in a separate section and signal transduction pathways (see supplemental Material, Table S5 and S6). Four expanded groups are characterized by the cytochrome P450 (PF00067) domain (Figure 4 C). Interestingly, mucoralean pathogens like *L. corymbifera* have been shown to be resistant to several antifungals including voriconazole [8,50] which could be explained by high copy numbers and isoforms of the target genes. Thus, gene duplication and expansion might be important for the success of *L. corymbifera* in human infections.

Strikingly, these domains (MFS transporters, HET and cytochrome P450) were also the dominant domains in genes which were localized in tandem duplications (Figure S3). Tandem duplicated genes were found to be present in 42 of the 66 gene expansion groups covering 38% of all genes in the expanded

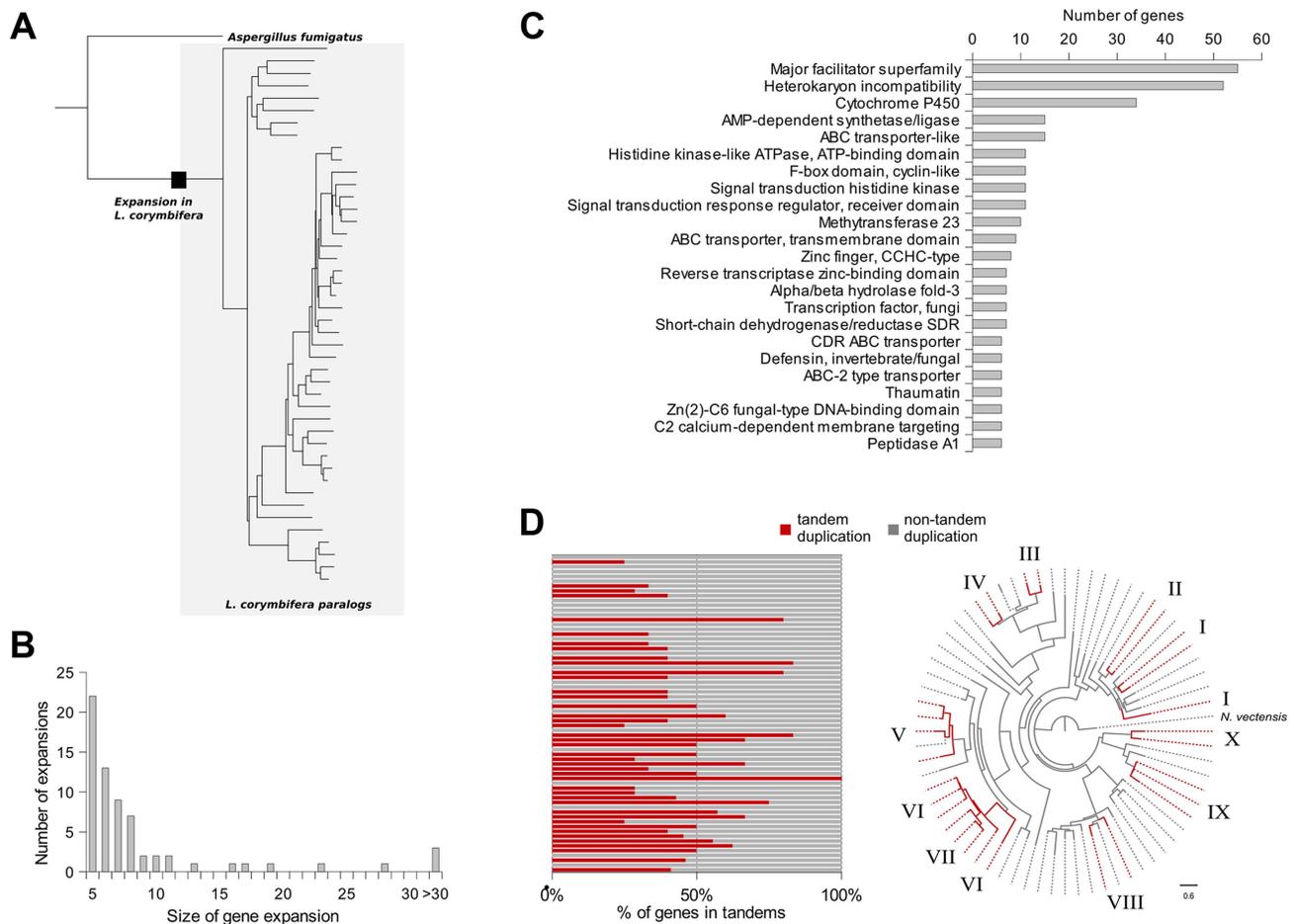


Figure 4. Gene expansions and tandem duplications found in *L. corymbifera*. (A) Tree representing an expansion of HET proteins in *L. corymbifera*. Branches enclosed in the grey shaded area represent paralogs of *L. corymbifera*. The black square represents the point where the expansion started. (B) Number and size of gene expansions in the *L. corymbifera* genome. (C) Main functional domains of gene expansions based on PFAM annotation. The numbers of genes with the different functional domains were combined if a domain was present in more than one expansion. (D) Proportion of genes within gene expansions which are arranged in tandem duplications. Each bar represents an expansion with the red part as the percentage of tandem duplicated genes (left). Clustering of tandem duplicated genes of cytochrome P450 genes in *L. corymbifera* (based on reconstruction with RaxML [130]). Red branches represent tandem duplicated genes. Numbers at the branch tips indicate different tandems. doi:10.1371/journal.pgen.1004496.g004

groups (Figure 4 D). However, additional smaller tandems were found which did not fit the criteria of gene expansions. A total of 701 genes are organized in such tandem repeats (Figure S3). In addition, duplicated genes were frequently found to be located on the same scaffold which may result from older tandem duplications.

Thus, tandem duplications and a high amount of gene retentions may give an additional explanation for the high amounts of duplicated genes in *L. corymbifera* comparable to the observations in plant genomes where segmental duplications (resulting from WGD) and tandem duplications play different roles in the enrichment of genes of several gene families [51,52]. Tandem duplication and the retention of duplicated genes would be an explanation for the severe differences in the size of gene families between mucoralean fungi with only 53% of gene families with the same size in *L. corymbifera* and *R. oryzae* (see Figure S3 C).

To investigate if the different gene copies may have different functions and thus may contribute to rapid adaptations to different environmental conditions we analysed the expression of tandem duplicated genes under infection-associated stress conditions (iron

depletion and hypoxia; see Table S7). Differential expression of at least one gene of the tandem clusters under at least one of the conditions was found for 71 tandems. Strikingly, only 7 tandems were co-regulated while in 64 cases expression of the copies was different including six cases were copies were antithetically regulated (Figure S4). These results are consistent with the hypothesis that the high prevalence and maintenance of duplicated genes leads to diversification of gene functions.

Alternative splicing

Duplicated genes can lead to the diversification of gene functions of the two copies which has been discussed in the section above. In addition, alternative splicing (AS) can increase the functional diversity. Gene prediction resulted in 841 alternative splicing events in a total of 683 genes (5.5% of total genes) comparable to the situation in *S. cerevisiae* [53]. Based on the analysis of the RNAseq data alternative splicing could be verified for 273 genes (2.2% of total genes) (Figure S5 A and Table S8). Alternative donor and acceptor are the dominant groups of alternative splicing events (>75% of the total events) which is similar to the situation in several higher eukaryotes [54] (Figure S5

A). Comparison of alternatively spliced genes with genes in tandem duplications and gene expansions showed that only 12 (4.4% of genes with AS) in these groups are also alternatively spliced. If AS occurs in tandem duplicated genes, it occurs in only one of the copies except in one case. This is in accordance with recent results in *S. cerevisiae* which show that duplicated genes can replace one alternatively spliced gene and that alternative splicing is often lost after gene duplication [55]. To test if AS plays a role in the stress adaptation of *L. corymbifera* we analysed the potential alteration in alternative splicing pattern during stress adaptation. Significant changes were only detected for 16 and 23 genes under iron depletion and hypoxia (<0.2% of the total genes), respectively (Figure S5 B). Based on the high incidence of gene duplication and the differential expression of the copies as well as the comparably low number of alternatively spliced genes, maintenance of duplicated genes seems to play a more important role for the generation of functionally distinct paralogs than alternative splicing.

Identification and expression of potential virulence factors under infection-related conditions

a) Iron uptake genes. Iron is an essential trace element for all organisms and plays a crucial role in fungal pathogenicity [56–58]. Elevated host iron levels are an important prerequisite for mucormycosis and the iron permease FTR1 has been shown to be crucial for virulence in *R. oryzae* [59–62]. The genome of *L. corymbifera* contains four copies of FTR1, of which three are located next to a multicopper oxidase and may share the same promoter. Both FTR1 and multicopper oxidase are important players of the reductive pathway and have been shown to form functional complexes in *C. albicans* [63]. Thus, co-expression of both genes and maintenance of their proximity may contribute to effective iron uptake and, hence, to the virulence of *L. corymbifera*. The higher copy number of FTR1 in *L. corymbifera* compared to *R. oryzae* (one copy) suggests more efficient employment of this pathway and, probably, optimisation of the different copies to different environmental conditions in *L. corymbifera*. To investigate the expression of iron-uptake genes under iron limited conditions we added the iron chelator bathophenanthrolinedisulfonic acid (BPS) to overnight cultures of *L. corymbifera* (200 μ M final concentration) and analysed gene expression via RNA sequencing (see Material and Methods for details). Consistent with this hypothesis only one of the copies of FTR1 (LCor01036.1) was up-regulated under iron limitation in the gene expression experiments while another copy (LCor06326.1) was moderately down-regulated (Figure 5 A). In addition, all multicopper oxidases which are co-localized with FTR1 were regulated in the same manner as the corresponding FTR1 gene and expressed at comparable levels (Figure 5 A). The two remaining copies (LCor00518.1, LCor04103.1) were expressed constitutively at low levels and may be either specific for other conditions or generally silenced. In addition to FTR1 and multicopper oxidases, ferric reductases are key elements in the reductive pathway of iron acquisition. Two of the three ferric reductases were up-regulated under iron limitation (LCor07115.1, LCor11373.1), whereas the third one was constitutively expressed (Figure 5 B).

Besides the role of the reductive pathway little is known about the iron uptake systems in mucoralean pathogens. Our analysis uncovered the presence of additional genes involved in iron uptake including zinc/iron permeases, heme oxygenases and siderophore transporters (Table 2). Heme utilization may contribute to growth within the host, since mucoralean pathogens rapidly invade blood vessels [3] and may use hemoglobin as iron source [60]. Accordingly, one of the heme oxygenases (LCor09772.1) was

strongly up-regulated under iron limitation (Figure 5 B). All available mucoralean genomes, including *L. corymbifera*, lack non-ribosomal peptide synthetases (NRPSs) and are therefore unable to produce hydroxamate siderophores. Instead they produce polycarboxylate siderophores (rhizoferrin), which have a much weaker binding activity compared to hydroxamate siderophores and are produced by direct fermentation [64,65]. In addition, mucoralean fungi are also able to utilize deferoxamine, a bacterial siderophore which is used as an iron-chelator in human therapy [66,67]. Interestingly, zygomycetous species have been shown to live in close relationship with bacteria, including cases containing bacterial endosymbionts indicating that xenosiderophores might play a role in the development of siderophore uptake systems [68,69]. However, *L. corymbifera* has also been shown to be a potent producer of siderophores itself [70]. Under iron-limitation the putative siderophore transporter of *L. corymbifera* (LCor01340.1) was up-regulated (Figure 5 B) supporting the role of siderophores in the iron acquisition of mucoralean fungi. Interestingly, based on the expression data we found an additional gene (LCor00410.1) that may be involved in the siderophore metabolism of *L. corymbifera*, containing functional domains which are typical for genes involved in regulation and synthesis of siderophores in bacteria (*lucA/lucC* PF04183 and *FhuF* PF06276). Thus, the gene may encode a novel candidate for a regulator of siderophore biosynthesis in mucoralean fungi.

Although fungi generally lack ferritin as intracellular iron storage, ferritin has been found in several mucoralean species [62,71]. Based on orthology searches, ferritin genes could be identified in all mucoralean genomes. In addition, domain search in *Spizellomyces punctatus* and *Allomyces macrogynus* (Origins of Multicellularity, BROAD) revealed also the presence of potential ferritin genes. Since these two groups represent the most basal fungal lineages, apart from the highly derived microsporidians [72], ferritin seems to have been lost in some microsporidians and the higher fungi. Interestingly, in higher fungi, the loss of ferritin coincides with the appearance of *sidA*, an important gene in hydroxamate siderophore production (Figure S6). Siderophores are known to serve as intracellular iron storages in several Asco- and Basidiomycota species [73,74]. Thus, a plausible hypothesis is that maintenance of ferritin was not necessary in derived fungi due to the gain of importance of siderophores. The expression of the two ferritin genes (LCor08103.1, LCor11038.1) was slightly decreased under iron-limitation ($\sim 1.6\times$, Fig. 5B). Nothing is known about the dynamics and functions of fungal ferritins. The slight decrease may be sufficient to stabilize free iron concentrations in the cytoplasm.

Only three transcription factors were up-regulated under iron depleted conditions (Figure S7). Interestingly, one of these genes (LCor08192.1) shows similarities to GATA type regulators which are known to be involved in the adaptation to iron limitation in higher fungal pathogens (e.g. *A. fumigatus* and *Histoplasma capsulatum*) [75,76]. Thus, this transcription factor may represent a key regulator in iron acquisition, and therefore an important virulence factor of *L. corymbifera*. A second transcription factor (LCor01236.1) resembles CRZ1, which is a calcineurin regulated TF and indicates a possible involvement of the calcineurin pathway in the adaption to iron limitation.

b) Secreted proteases. Besides the iron uptake systems, hydrolytic enzymes like proteases are known as important virulence factors in fungal pathogens, e. g., *R. microsporus* and *C. albicans* [77–79]. In addition, gene expansion of secreted proteases was observed in the *R. oryzae* genome [28]. The genome of *L. corymbifera* contains a total of 413 predicted proteases representing 3.3% of all genes comparable to the situation in *R.*

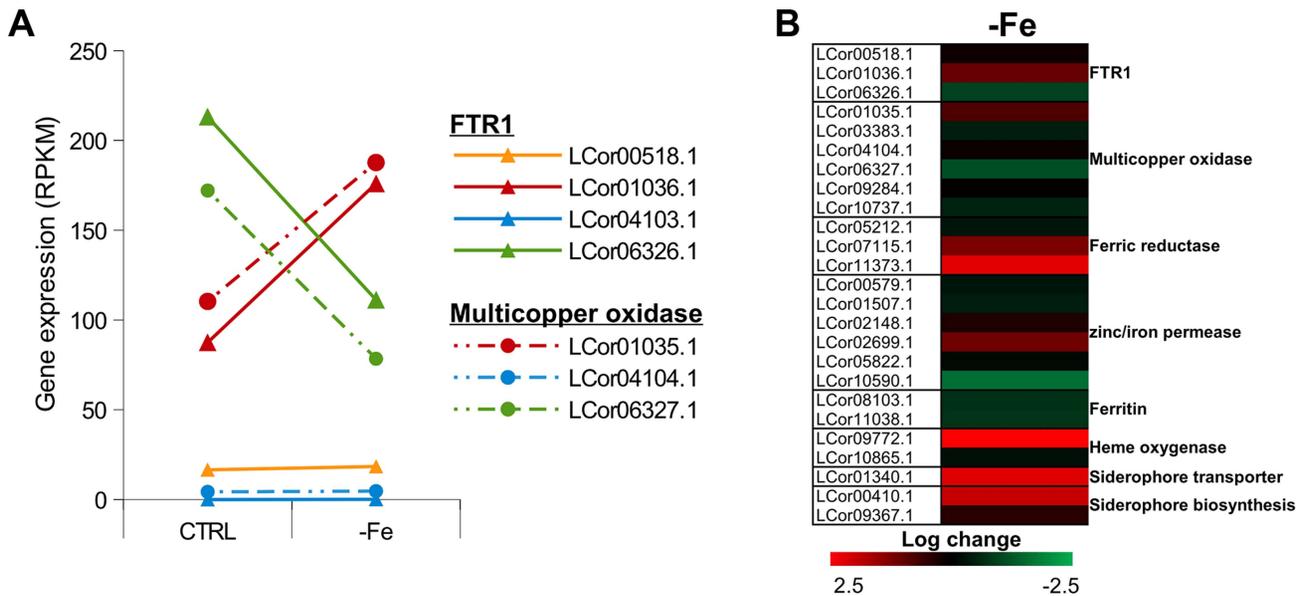


Figure 5. Expression of iron uptake genes under iron limited conditions. (A) Expression levels of FTR1 domain genes and their corresponding multicopper oxidases under standard conditions and iron limitation. (B) Heat map showing the regulation of iron uptake genes under iron-limited conditions.

doi:10.1371/journal.pgen.1004496.g005

oryzae which contains 630 proteases (3.6% of all genes). However, the number of secreted proteases differs between the two species, moreover, in *L. corymbifera* the relative amount of secreted proteases is nearly twice as high as in *R. oryzae*: in *L. corymbifera* 13% (53) of the proteases are predicted to be secreted while in *R. oryzae* this number reaches only 7% (44). The most important classes of secreted proteases are serine and aspartate proteases representing 55% and 36% of total secreted proteases, respectively (Table 3). Comparison of secreted aspartic proteases (SAP) of *R. oryzae* revealed an enrichment of SAPs compared to other fungal genomes [28]. However, the number of SAPs is comparable in *L. corymbifera* (24) and *R. oryzae* (28) indicating that the presence of a higher number of these enzymes is a general feature of mucoralean pathogens. Several secreted proteases were activated under infection-related stress conditions (iron depletion, hypoxia). While iron depletion affected mainly the expression of aspartic and serine proteases, hypoxic conditions induced the expression of serine-, metallo- and some aspartic proteases.

c) Transcription factors. The *Lichtheimia* genome encodes 768 putative transcription factors (TFs) representing 6.2% of total genes. This amount is comparable to the situation in *R. oryzae*

(6.4% of total genes) but higher than the average content of TFs in other fungi (4.5%) [80]. Basic BLASTp analyses showed that 37 of the TFs (4.8% of total TFs) are specific for *L. corymbifera*.

The TFs were assigned to 53 families of DNA-binding domains (based on the InterProScan predictions). The great majority of these families have been described previously in fungal species [81]. However, 4 families have not been found in true fungal species before (putative representatives of 2 families, PF01167 (Tubby) and PF02319 (TDP), were predicted in microsporidia; the other 2 were described in plants and bacteria). Of these “new” TFs one (PF03106, DNA-binding WRKY domain), represented in 7 *Lichtheimia* proteins, has been previously described only in plants. Interestingly, in plants these TFs are numerous and have diverse functions, including pathogen defense [82]. In *L. corymbifera* we found three of the seven members differentially regulated under hypoxic conditions (down: LCor09690.1; up: LCor02851.1, LCor08197.1) indicating a function in the stress response of this species (Figure S7).

Of 6 fungal-specific TF families [81], 2 families are not predicted in *L. corymbifera* genome, namely PF04769 (mating-type protein MAT a1) and PF02292 (APSES domain). The lack of

Table 2. Iron uptake genes in the *L. corymbifera* genome.

Pathway	Iron uptake gene	Number of genes
Reductive pathway	FTR1	4
	multicopper oxidase	8
	ferric reductase	3
Low affinity iron uptake	zinc/iron permease	6
Siderophore uptake	siderophore transporter	1
Heme utilization	heme oxygenase	2
Iron storage	ferritin	2

doi:10.1371/journal.pgen.1004496.t002

Table 3. Protease families in the *L. corymbifera* genome.

	Proteases (% of total proteases)	Secreted (% of secreted proteases)
Aspartate	60 (14.5)	19 (35.8)
Cysteine	68 (16.5)	2 (3.8)
Metallo	146 (35.4)	3 (5.6)
Serine	124 (30)	29 (54.7)
Threonine	14 (3.4)	0
Unknown	1 (0.2)	0
Total	413	53

doi:10.1371/journal.pgen.1004496.t003

MAT gene is expected because mating in Mucorales is regulated via sex plus and sex minus HMG transcription factors [83]. The absence of APSES domain may be compensated by the APSES-type DNA binding domain PF04383.

It is also noteworthy that the traditional proportion of the Zn fingers Zn₂Cys₆ (Zn cluster) and Cys₂His₂ is inverted in *Lichtheimia*. In all fungi observed so far, the Zn clusters are more abundant than C₂H₂ TFs, in fact they are normally the most numerous in the fungal genomes. In *Lichtheimia*, on the contrary, the number of C₂H₂ Zn fingers is nearly 1.5 times larger than the number of Zn clusters.

Comparative phylome-based analysis reveals several expanded TF families in *L. corymbifera* including MADS box TFs with 11 representatives instead of the usual 1–4 members. MADS box genes are known to play a role in a variety of functions (e.g. cell cycle, stress response, development [84]). Presumably, the expansion of the MADS box genes in *Lichtheimia* was accompanied by the delegation of some functions from other TFs or even neofunctionalization. The functional basis for such expansion as well as the potential roles of these TFs cannot be elucidated from their primary structure, because MADS box genes are not conserved except for the MADS domains. But at the expression level, we could find significant up-regulation of two of the 11 MADS box TFs (LCor03918.1, LCor08105.1). Thus, the copies do not seem to have completely overlapping functions.

Analysis of the phylomes allowed us to detect another exciting expansion, which is evidently characteristic for all Mucorales: the duplication of TBP, TATA binding protein (PF00352). As it has been recently shown for higher eukaryotes, core promoter recognition factors can be involved in modulating gene- and cell-type-specific programs of transcription, such as tissue differentiation, development, etc. [85]. These new functions are associated with a gene duplication of the TBP, resulting in TRF2 (and other) factors, which are highly similar to TBP but do not bind the TATA box. In fungi, the event of TBP duplication is exceptionally rare. A survey of all so far sequenced genomes revealed only 4 examples of such duplication: 3 in Ascomycetes/Sordariomycetes (*Chaetomium globosum*, *Grosmannia clavigera* and *Podospora anserina*) and 1 in the basidiomycete *Laccaria bicolor*. In contrast, in Mucorales all 4 considered species possess 2 copies of the TBP gene. It can be supposed that the duplicated TBP-like factors may play an additional role in condition-specific responses and thus may be of interest as potential virulence factors.

Temperature tolerance is an essential prerequisite for the infection of warm-blooded animals and was shown to be connected to the virulence of *Lichtheimia* species [25,86]. The genome was surveyed for the presence of heat shock transcription factors (HSF). The total number of these TFs in *L. corymbifera*

genome is 24, which is the highest number among all so far investigated fungi. This is in accordance with the known tolerance of *L. corymbifera* to high temperatures [13,14]. However, it seems that this family expansion is not a specific trait of *Lichtheimia* but is characteristic for all Mucorales. Interestingly, HSF genes were also up-regulated under hypoxic conditions indicating additional functions of the different members of the HSF family in the response to different stresses and growth conditions.

It is curious that additionally to the abundant heat shock factors also a cold shock TF (PF00313) was found, which was not previously described in fungi. This can explain why *Lichtheimia*, although it does not grow at low temperatures, can tolerate cold as it was shown to survive periods of more than 5,000 years in ice [87].

The importance of the *L. corymbifera* genome for studying the infection biology of mucoralean pathogens: Concluding remarks

The genome of *L. corymbifera* represents the first insight into the genome structure of basal mucoralean pathogens. Despite the growing recognition of Mucorales as life-threatening clinically important human pathogens, little is known about the virulence traits of these fungi. The high dissimilarity between *L. corymbifera* and the other sequenced mucoralean pathogens *R. oryzae* and *M. circinelloides* in both evolutionary and functional sense underlines the importance of additional genome projects.

This study revealed a high proportion of duplicated and expanded genes in the *L. corymbifera* genome comparable to the situation in *R. oryzae*. However, clear evidence for a WGD can be detected only for *R. oryzae*, but not for *L. corymbifera* indicating that additional mechanisms contribute to the higher incidence of duplicated genes in mucoralean fungi. Tandem repeats seem to be an important source for gene duplication in *L. corymbifera* and may explain the rapid development of lineage-specific gene duplication and expansion in mucoralean fungi. Several species-specific gene duplications point at potential virulence traits including iron uptake genes, hydrolytic enzymes and genes which may contribute to resistance against antifungal agents like azoles (cytochrome P450 gene expansion). In contrast, alternative splicing does not seem to play an important role in the generation of orthologs and the adaptation to stress conditions.

Based on these results, we postulate a relationship between genome fluidity by the generation and retention of additional gene copies and dynamics of adaptation to new environments. Higher genome flexibility results in a higher likelihood for a saprobic zygomycete to become a pathogen.

In addition we were able to shed light on the genes involved in iron uptake, which is a crucial step for virulence and thus for the

development of an infection. We could identify additional genes which might be involved in iron-uptake besides the known virulence factor FTR1 *L. corymbifera* including transcription factors, siderophore transporters and a potential regulator involved in siderophore biosynthesis that has not been described in mucoralean fungi.

Our data represent a valuable resource for future research and the understanding of infection-associated mechanisms of mucoralean pathogens.

Materials and Methods

Genome sequencing and assembly

A combination of Illumina and 454 sequencing was used for the *L. corymbifera* genome. A shotgun library and an 8 kb paired-end library were created and sequenced on a half plate on a Roche GS FLX Titanium each resulting in 1,168,226 shotgun reads (505,023,982 nt) and 519,989 paired-end reads (76,603,029 nt). In addition, a standard paired-end read library was prepared and sequenced in one channel Illumina HiSeq2000 (100 bp paired-end reads) resulting in 264,907,616 raw reads (26,490,761,600 nt) and 12,614,650 filtered and downsampled reads (1,261,465,000 nt). The 454 reads were separately assembled using Newbler (454 Life Sciences) and Mira [88] and both assemblies were unified using minimus2 [89]. The Illumina reads were used to solve homopolymeric regions using Nsoni (<http://bioinformatics.net.au/software>). This approach resulted in a total of 1,214 contigs (\geq 500 nt) with a total of 41,405,106 nt and a N50 of 66.718 nt. Finally, the contigs were mapped on Newbler predicted scaffolds using MUMmer [90] resulting in 209 scaffolds with a total length of 33.6 Mb (for statistics refer to Table 1). The raw DNA-seq reads and the resulting genome assembly is available at EMBL under the study accession number PRJEB3978 (<http://www.ebi.ac.uk/ena/data/view/PRJEB3978>).

Detection of transposable elements

Scaffolds of *L. corymbifera* were searched for repeats by Repbase and the server version of Censor [91,92] (<http://www.girinst.org/censor/index.php>) using the eukaryotic repeat database.

K-mer analysis

The analysis was performed on the Illumina reads with an algorithm described in the potato genome paper [31]. The algorithm was used to write a custom perl program. Based on the fastq data of the Illumina reads k-mers of 41, 59, 69, and 79 nt were detected and analyzed. Component estimation was done manually in R.

Non-coding RNA prediction, synteny and phylogeny

A local version of tRNAscan-SE v.1.23 [93] with parameters –omlfrf was used for the detection of tRNAs. Output files are supported in the supplemental material (<http://www.rna.uni-jena.de/supplements/lichtheimia/index.html>). With RNAmmer -S euk -m lsu,ssu,tsu -gff (v.2.1) [94] rRNAs were detected. The 1973 ncRNA classes currently available at RFAM (v.10.1) [95] were downloaded for homologous search. These classes were predicted with (I) BLAST (v.2.2.25) [96] with an E-value $<10^{-4}$ (II) with infernal [97] using covariance models from RFAM and (III) by hand as indicated in main text. Genes discovered in the reads only were found with rnabob [98] in combination with various programs of the RNAViennaPackage v.2.0.2 (<http://www.tbi.univie.ac.at/~ivo/RNA/>). All ncRNA genes are available at the supplemental material in gff and fasta format (<http://www.rna.uni-jena.de/supplements/lichtheimia/index.html>).

Additionally, sequence-structure-alignments for each RFAM-ncRNA class in stockholm format are provided. Motif search in promoter regions of polymerase III transcripts was performed with MEME (v.4.8.1) [99], rnabob and by hand. Synteny analysis: for all of our identified ncRNA positions in *L. corymbifera* and *R. oryzae*, five direct upstream and downstream located genes and their function were extracted, according to protein-annotation files. Pairwise alignments of syntenic proteins with a) -p blastn and b) -p tblastn and a minimum E-value of $E < 10^{-4}$ were performed. For ncRNA-phylogeny reconstruction the best scored ncRNA per ncRNA family was joined, which was identified in all species, except 18S and 28S rRNA, and *S. pombe* used as outgroup. A multiple alignment was created by Mafft with the L-INS-i method, 1000 iterations as module in the EPoS framework for phylogenetic analysis [100]. Out of this alignment we constructed a Neighbour Joining Tree (Kimura correction model, 1000 bootstrap replicates) and Mr. Bayes (v.3.1.2; two runs with each four chains and 5,000,000 generations).

Prediction of protein-coding genes and functional annotation

Evidence-driven gene prediction was performed using AUGUSTUS v2.7 [37] using the gene models from *Rhizopus oryzae* prediction was supported by the incorporate pooled Illumina RNA-seq data from three biological replicates of three different physiological conditions (control, hypoxia, iron depletion) sequenced on Illumina HiSeq 2000. After the raw RNA-Seq data were quality trimmed- using btrim [101], the data were pooled and mapped using the splice-junction mapper tophat2 [102]. From this mapping data the AUGUSTUS protocol (<http://bioinf.uni-greifswald.de/bioinf/wiki/pmwiki.php?n=IncorporatingRNAseq>. Tophat) was followed to create hints for gene structures in an iterative manner. Finally, the hints were incorporated during the AUGUSTUS prediction based on the *Lichtheimia* genome using the metaparameters of *R. oryzae*. For functional annotation predicted protein-coding genes were analyzed by BLASTp in BLAST2GO [38] with a minimum E-value of $E \leq 10^{-25}$ and a HSP length cut-off of 33 amino acids. Conserved domains were identified using the InterProScan function of BLAST2GO and GO mapping was performed based on the BLAST and InterProScan results. Genome annotations are available at the ENA under the study accession number PRJEB3978 (<http://www.ebi.ac.uk/ena/data/view/PRJEB3978>).

Detection of differentially expressed genes under infection-related conditions

Paired-end RNA-seq data for three biological replicates of two infection-associated conditions (i.e., iron-depletion and hypoxia) and a control treatment was obtained.

L. corymbifera was grown on SUP agar [103] plates for 7 days at 37°C. Spores were washed off with sterile PBS, washed with PBS and counted using a Thoma chamber. Erlenmeyer flasks (500 ml) containing 100 ml of chemical defined medium (1.7 g/l YNB w/o amino acids and ammonium sulphate, 20 g/l Glucose, 5 g/l ammonium sulphate, 50 mg/l arginine, 80 mg/l aspartic acid, 20 mg/l histidine, 50 mg/l isoleucine, 100 mg/l leucine, 50 mg/l lysine, 20 mg/l methionine, 50 mg/l phenylalanine, 100 mg/l threonine, 50 mg/l tryptophane, 50 mg/l tyrosine, 20 mg/l valine) [60] were inoculated with 10^7 spores and grown for 16 h at 37°C under shaking. Afterwards (i) cultures were grown for additional 2 h under these conditions, (ii) the iron chelator bathophenanthroline-disulfonic acid (BPS, Sigma) was added to a final concentration of

200 μm and cultures were incubated for additional 2 h under previous conditions or (iii) cultures were subjected to hypoxic conditions (1% oxygen, 5% CO_2) and incubated for 2 h at 37°C under shaking. The mycelium was separated from the medium using a miracloth filter (Millipore) and immediately frozen in liquid nitrogen. For RNA isolation the mycelium was grounded using mortar and pestle under liquid nitrogen and total RNA was isolated using the RNeasy Plant kit (Qiagen) according to the manufacturer's instructions.

Sequencing was performed using Illumina HiSeq 2000. Raw reads were quality-filtered using btrim [101] and mapped to the genome using tophat2 [102] (parameters: `-no-discordant -no-mixed -b2-very-sensitive -max-intron-length 5000`). Differentially expressed genes were identified with EdgeR [104] which also adjusted obtained p-Values for multiple testing. Transcripts with an absolute fold-change ≥ 2 and an adjusted p-Value ≤ 0.01 were considered differentially expressed. Results are available in Table S7.

Phylome reconstruction

The phylome, meaning the complete collection of phylogenetic trees for each gene in a genome, was reconstructed for the genome of *L. corymbifera*. 24 other fungal species were included in the reconstruction. A rough draft of the proteome of *Mortierella alpina* ([26]; PUBMED ID:22174787) was predicted using AUGUSTUS [32] due to the lack of a publicly available proteome. The phylome was reconstructed using an automated pipeline previously described in [39]. Briefly, for each protein in the *L. corymbifera* genome a Smith-Waterman search was performed against the fungal proteome database. Results were filtered using an e-value cut-off $E < 1e^{-5}$ and a continuous overlapping region of 0.5. At most 150 homologous sequences for each protein were accepted. Homologous sequences were then aligned using three different programs: MUSCLE v3.8 [105], MAFFT v6.712b [106], and kalign (<http://www.biomedcentral.com/1471-2105/6/298/>). Alignments were performed in forward and reverse direction (i.e. using the Head or Tail approach [107]), and the 6 resulting alignments were combined with M-COFFEE [108]. This combined alignment was trimmed with trimAl v1.3 [109] (consistency-score cut-off 0.1667, gap-score cut-off 0.9). Trees were reconstructed using the best-fitting evolutionary model. The selection of the model best fitting each alignment was performed as follows: a Neighbour Joining (NJ) tree was reconstructed as implemented in BioNJ [110]; the likelihood of this topology was computed, allowing branch-length optimization, using 7 different models (JTT, LG, WAG, Blosum62, MtREV, VT and Dayhoff), as implemented in PhyML v3.0 [111]; the model best fitting the data, as determined by the AIC criterion [112], was used to derive ML trees. Four rate categories were used and invariant positions were inferred from the data. Branch support was computed using an aLRT (approximate likelihood ratio test) based on a chi-square distribution. Resulting trees and alignments are stored in phylomeDB [39] (<http://phylomedb.org>), with the phylomeID 245. Trees were scanned using ETE v2 [46].

Orthology prediction

Orthologs between *L. corymbifera* and the other species included in the phylome were based on phylogenies obtained during phylome reconstruction. A species-overlap algorithm, as implemented in ETE v2 [46], was used to infer orthology and paralogy relationships. Briefly the algorithm decides whether a node in a tree is a speciation of a duplication node depending on the overlap of the species branching from the node. Overlap

between those species will indicate a duplication node. Otherwise a speciation node will be considered.

Species tree reconstruction

The species tree was built using a concatenation method. 58 single-copy proteins that appeared in at least 21 of the 25 genomes were selected. After concatenation, the alignment was trimmed using trimAl [109]. Columns with more than 50% of gaps were removed. A conservation score of 50% of the alignment was used. The final alignment contained 46,793 positions. The tree was reconstructed using phyML [111]. LG model [113] was selected and a 4-categories GAMMA distribution was used. Bootstrap was obtained by creating 100 random sequences using SeqBoot from the phylib package. A tree was then reconstructed for each sequence and the consensus tree was inferred using phylib. All the nodes in the species tree had a bootstrap of 100. Additionally a species tree based on the super-tree reconstruction program DupTree [42] was reconstructed. The input contained the 9,478 trees obtained during phylome reconstruction. Both species trees showed a similar topology. The only difference pertained to the position of *S. pombe*. In the concatenated tree it appeared grouped with *S. cerevisiae* while in the super-tree it appeared in its correct position at the base of Ascomycota. This difference was collapsed into a multifurcation for the tree in figure 1.

Detection of conserved regions

For the detection of conserved regions, all genomes were modeled as strings of integers. BLAST analyses [96] were performed for all proteins in the four mucoralean genomes all-against-all, with an E-value threshold of 0.1. Homology families IDs were assigned to the protein-coding genes using GhostFam [114] with default parameters. Genomes were transformed into strings of gene IDs, which were then used as input for the reference gene cluster implementation in Gecko2 [43,44]. The two parameters for the algorithm were the minimum size of the reference cluster/hypothetical conserved region “s” and the maximal distance “ δ ” (insertion or deletion of a gene). For every hypothetical gene cluster larger than s on the reference genome, all other genomes were tested for approximate occurrences of this reference gene cluster. The *L. corymbifera* genome was used as a reference genome and searched for gene clusters with parameters $s = 3$ (minimum size of the reference gene cluster) and $\delta = 0$ (number of insertions and deletions), $s = 4/\delta = 1$, $s = 5/\delta = 2$, $s = 6/\delta = 3$ and $s = 7/\delta = 4$. Results of the different filter settings were combined and overlapping clusters were eliminated. Local rearrangements and duplications within the cluster occurrences were not punished. All regions that had approximate occurrences in at least one other genome were reported. If multiple occurrences did intersect, only the best scoring one was reported.

Detection of duplicated regions (segmental duplications)

To detect duplicated regions in the mucoralean species, each genome was analysed individually by using the single contigs as reference. As for the detection of conserved regions, the same homology assignment and parameters of $s = 5$ and $\delta = 2$ were used. All regions with approximate occurrences in at least one other contig or the reference contig were reported, unless they intersected.

Detection of tandem duplications

Tandem duplications were defined by at least two genes assigned to the same GhostFam gene family and a maximum of three genes between the copies.

Prediction of alternative splicing

Predicted transcripts of the genomes were separated in alternative splicing events by Astalavista [24]. Events of predicted transcripts that contain splice-junctions have been confirmed by the number of split-mappings that confirm each of the exon-exon junctions (Table S8). For a read to support a splice-junction, the left part of the read was required to be included in one exon, and the right part had to be included in the other exon of a splice junction, with the first/last position before/after the split matching exactly the position of the predicted intron.

Genome resources

Genome data of *Aspergillus fumigatus* [115], *Aspergillus nidulans* [116], *Batrachochytrium dendrobatidis*, *Cryptococcus neoformans*, *Encephalitozoon cuniculi* [117], *Rhizopus oryzae* [28], *Paracoccidioides brasiliensis*, *Schizosaccharomyces pombe* [118], *Nosema ceranae* [119], *Nematocida parisii* [120], *Puccinia graminis* [121], *Ustilago maydis* and *Coprinus cinerea* [122] are genome sequencing projects of the Broad Institute of Harvard and MIT (<http://www.broadinstitute.org/>) (see Table S9 for detailed citations). *Phycomyces blakesleeanus*, *Phanerochaete chrysosporium* [123], *Laccaria bicolor* [124], *Mucor circinelloides*, *Nematostella vectensis* [125], *Monosiga brevicollis* [126] and *Serpula lacrymans* [127] genomic data were obtained from Joint Genome Institute (JGI). These sequence data were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community. The genomes of *Homolophyctis polyrhiza* [128] and *Mortierella alpina* [26] were obtained from Genbank (*Hp*: PRJNA68115; *Ma*: PRJNA41211). The *Neurospora crassa* genome [129] was obtained from UniProt reference genomes. The *Saccharomyces cerevisiae* genome was obtained from *Saccharomyces* Genome database (SGD) (see Table S9) [53].

Supporting Information

Figure S1 K-mer frequency distribution for *Lichtheimia corymbifera*. The k-mer frequency distribution (black line) was calculated for all k-mers of length of 59, i.e. for all possible 59-mers derived from the original Illumina/Solexa reads. The number of k-mers (y-axis) is plotted against the frequency at which they occur (x-axis). The distribution shows a main peak (shaded in light gray) and a steep rise to the left (shaded in dark gray). This left-most rise of k-mers at lower frequencies represents mostly k-mers with randomly occurring sequencing errors. The main peak represents k-mers derived from (putatively) correct sequencing reads. This main peak can be dissected into three normal distributions (red, blue and orange lines) the sum of which (green line) matches the observed distribution (black line). The three component distributions represent the ‘homozygous’ part of the genome (blue line, major component), the ‘heterozygous’ part of the genome (red line), and most likely some repeat regions that make up a minor proportion of the observed k-mers (orange line). Component estimation was done manually in R. The component distributions have the same variance (21), but different means (blue 116, red 65, orange 165) and proportions (blue 94%, red 4%, orange 2%). (TIFF)

Figure S2 Structure of spliceosomal RNAs and ncRNA based phylogeny. (A) The last stem (IV) of *Lichtheimia* U11 snRNA is extended in comparison to other U11 snRNAs and to U1 snRNA. U2 snRNA folds into an expected secondary structure. In contrast, U12 snRNA shows an extended stem II, and misses the third stem (III). Stem IV/V is much shorter as in other known U12 snRNAs.

U5 snRNA is used by both spliceosomes, with the general eukaryotic secondary structure. 2D structures were computed using RNAfold (RNA Vienna Package). Boxes indicate sm binding sites. Phylogeny of *L. corymbifera*, *M. circinelloides*, *P. blakesleeanus*, *R. oryzae* and *S. pombe* (outgroup) based on ncRNAs (except 18S and 28S rRNA). Alignment computed via Mafft L-INS-i with 1000 generations; Tree construction via (B) Neighbour Joining: Kimura: 1000 bootstrap replicates and (C) Mr. Bayes: two runs with each four chains and 5,000,000 generations. (TIFF)

Figure S3 Tandem duplications in *L. corymbifera*. (A) Number and size of tandem duplication in the *L. corymbifera* genome. (B) Functional classes of genes in tandem duplications based on PFAM annotation. Asterisk indicates classes which are enriched in tandem duplications (Fisher test, $P < 0.05$). (C) Gene family size comparison of *L. corymbifera* and *R. oryzae*. Gene families are indicated as larger in *Lichtheimia* ($L > R$), smaller in *Lichtheimia* ($L < R$) or as large as in *Rhizopus* ($L = R$). (TIFF)

Figure S4 Expression of tandem duplicated genes. Tandem duplicated genes were analysed based on the RNA-seq data. Tandems were regarded as (i) not regulated if no copy in the cluster was up/down-regulated under the tested conditions, (ii) co-regulated if all copies in the clusters were up/down-regulated under at least one of the conditions, (iii) not co-regulated if one of the copies was differently regulated than the other(s), (iv) antithetically regulated if one copy was up- and the other down-regulated. Genes were regarded as differentially regulated if there was a two-fold change of expression and $P < 0.01$ (edgeR). (TIFF)

Figure S5 Alternative splicing in *L. corymbifera*. (A) Number and proportion of different classes of alternative splicing events based in evidence driven gene prediction (outer ring) and confirmed events (inner ring). (B) Proportion of AS genes where AS patterns were changed under stress conditions compared to control. (TIFF)

Figure S6 Distribution of genes involved in iron uptake within the fungal kingdom. Orthologs of iron uptake genes were identified using the phylogeny of *L. corymbifera* (indicated in blue). If no ortholog was found BLASTp analysis was performed using the *L. corymbifera* protein sequence and an E-value $E \leq 10^{-10}$ (indicated in pink). Intracellular iron storages besides ferritin are indicated as ‘s’ (siderophores) or ‘v’ (vacuolar) according to previous results (1 Silva et al., 2 Haas et al.). The presence of a sidA ortholog is indicated as “+”, the absence as “-” according to previous results (1 Silva et al., 2 Haas et al.). (TIFF)

Figure S7 Differential expression of transcription factors under iron depletion and hypoxia. Bar charts on top represent TFs grouped according to their functional domains (domain combinations). Up- and down-regulated genes are indicated in red and green respectively. The bar chart on the bottom shows the total amount of TFs regulated under the conditions. (TIFF)

Table S1 Sequencing statistic of the *L. corymbifera* genome and transcriptome. (PDF)

Table S2 Transposable and repetitive elements in the *L. corymbifera* genome. (PDF)

Table S3 Overview of ncRNAs found in basal fungi. op – Close/Part to Operon; pg – Pseudogene; Ror – *R. oryzae*; Lco – *L. corymbifera*; ? – candidate. (PDF)

Table S4 Sequencing and mapping statistics of RNA sequencing. (XLS)

Table S5 Signalling pathway components included in the study and orthologues identified in *L. corymbifera*. (PDF)

Table S6 Classification and gene IDs of putative protein phosphatases in the *L. corymbifera* genome. (PDF)

Table S7 RNA-Seq mapping and differentially expressed genes. (XLS)

Table S8 Potential alternatively spliced genes in *L. corymbifera* and confirmation of alternative transcripts by RNA-Seq data. (XLSX)

Table S9 Genomes used in this study. (PDF)

Acknowledgments

The authors want to acknowledge Joint Genome Institute (JGI) and Broad Institute of Harvard and MIT for the permission for the usage of their genome data (see Genome resources in Material and Methods, and Table S9). The *Rhizopus oryzae* gene model was publicly provided by Jason Stajich (UC Riverside). Thanks to Katharina Hoff (University Greifswald) for her support using the Augustus gene predictor.

Author Contributions

Conceived and designed the experiments: VUS SWi ES MMH FH MS MN VV JL IDJ MM AAB TG SB KV. Performed the experiments: VUS SWi ES MMH FH MS MN VV JL IDJ KK KR SWe. Analyzed the data: VUS SWi ES MMH FH MS MN VV JL IDJ KR SWe MM TG SB KV. Contributed reagents/materials/analysis tools: SWi ES FH MS MN VV JL IDJ KR SWe MM AAB TG SB KV. Wrote the paper: VUS SWi ES MMH FH MS MN VV MM TG SB KV. Designed the software used in analysis: SB MS SWi KR MM FH JL MMH TG.

References

- Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycol Res* 111: 509–547. doi:10.1016/j.mycres.2007.03.004.
- Roden MM, Zautis TE, Buchanan WL, Knudsen T a, Sarkisova T a, et al. (2005) Epidemiology and outcome of zygomycosis: a review of 929 reported cases. *Clin Infect Dis* 41: 634–653. doi:10.1086/432579.
- Sugar A M (1992) Mucormycosis. *Clin Infect Dis* 14 Suppl 1: S126–9.
- Ribes JA, Vanover-Sams CL, Baker DJ (2000) Zygomycetes in human disease. *Clin Microbiol Rev* 13: 236–301.
- Chakrabarti A, Das A, Mandal J, Shivaprakash MR, George VK, et al. (2006) The rising trend of invasive zygomycosis in patients with uncontrolled diabetes mellitus. *Med Mycol* 44: 335–342. doi:10.1080/13693780500464930.
- Cornely OA, Vehreschild JJ, Rüping MJGT (2009) Current experience in treating invasive zygomycosis with posaconazole treatment options for zygomycosis. *Clin Microbiol* 15: 77–81.
- Skiada A, Pagano L, Groll A, Zimmerli S, Dupont B, et al. (2011) Zygomycosis in Europe: analysis of 230 cases accrued by the registry of the European Confederation of Medical Mycology (ECMM) Working Group on Zygomycosis between 2005 and 2007. *Clin Microbiol Infect* 17: 1859–1867.
- Vitale RG, de Hoog GS, Schwarz P, Dannaoui E, Deng S, et al. (2012) Antifungal susceptibility and phylogeny of opportunistic members of the order mucorales. *J Clin Microbiol* 50: 66–75. doi:10.1128/JCM.06133-11.
- Xess I, Mohapatra S, Shivaprakash MR, Chakrabarti A, Benny GL, et al. (2012) Evidence implicating *Thamnostylum lucknowense* as an etiological agent of rhino-orbital mucormycosis. *J Clin Microbiol* 50: 1491–1494. doi:10.1128/JCM.06611-11.
- Lanternier F, Dannaoui E, Morizot G, Elie C, Huerre M, et al. (2012) A global analysis of mucormycosis in France: The RetroZygo Study (2005–2007). *Clin Infect Dis* 54: 35–43. doi:10.1093/cid/cir880.
- Alvarez E, Sutton D a, Cano J, Fothergill a W, Stehigel a, et al. (2009) Spectrum of zygomycete species identified in clinically significant specimens in the United States. *J Clin Microbiol* 47: 1650–1656. doi:10.1128/JCM.00036-09.
- Tieghem, Van P (1876) Troisième mémoire sur les Mucorinées. *Ann des Sci Nat Bot* 4: 312–399.
- Hoffmann K, Discher S, Voigt K (2007) Revision of the genus *Absidia* (Mucorales, Zygomycetes) based on physiological, phylogenetic, and morphological characters; thermotolerant *Absidia* spp. form a coherent group, *Mycocladaceae* fam. nov. *Mycol Res* 111: 1169–1183. doi:10.1016/j.mycres.2007.07.002.
- Alastrucy-Izquierdo A, Hoffmann K, de Hoog GS, Rodriguez-Tudela JL, Voigt K, et al. (2010) Species recognition and clinical relevance of the zygomycetous genus *Lichtheimia* (syn. *Absidia* pro parte, *Mycocladus*). *J Clin Microbiol* 48: 2154–2170. doi:10.1128/JCM.01744-09.
- De Hoog GS, Guarro J, Gene J, Figueras M (2000) Atlas of Clinical Fungi. 2nd ed. Centraalbureau voor Schimmelcultures. p. 58–114.
- Haidle CW, Storck R (1966) Control of dimorphism in *Mucor rouxii*. *J Bacteriol* 92: 1236–1244.
- Cooper BH (1987) A case of pseudoparasitoidomycosis: Detection of the yeast phase of *Mucor circinelloides* in a clinical specimen. *Abstract. Mycopathologia* 97: 189–193.
- Hesseltine CW, Featherston C (1985) Anaerobic growth of molds isolated from fermentation starters used for foods in Asian countries. *Mycologia* 77: 390–400.
- Sun H-Y, Aguado JM, Bonatti H, Forrest G, Gupta KL, et al. (2009) Pulmonary zygomycosis in solid organ transplant recipients in the current era. *Am J Transplant* 9: 2166–2171. doi:10.1111/j.1600-6143.2009.02754.x.
- Bellanger A-P, Reboux G, Botterel F, Candido C, Roussel S, et al. (2010) New evidence of the involvement of *Lichtheimia corymbifera* in farmer's lung disease. *Med Mycol* 48: 981–987. doi:10.3109/13693781003713711.
- Copetti MV, Iamanaka BT, Frisvad JC, Pereira JL, Taniwaki MH (2011) Mycobiota of cocoa: from farm to chocolate. *Food Microbiol* 28: 1499–1504. doi:10.1016/j.fm.2011.08.005.
- Mphande FA, Siame BA, Taylor JE (2004) Fungi, aflatoxins, and cyclopiazonic acid associated with peanut retailing in Botswana. *J Food Prot* 67: 96–102.
- Baffi MA, Romo-Sánchez S, Ubeda-Iranzo J, Briones-Pérez AI (2012) Fungi isolated from olive ecosystems and screening of their potential biotechnological use. *N Biotechnol* 29: 451–456. doi:10.1016/j.nbt.2011.05.004.
- Hong S, Kim D, Lee M, Back S, Kwon S, et al. (2012) Zygomycota associated with traditional meju, a fermented soybean starting material for soy sauce and soybean paste. *J Microbiol* 50: 386–393. doi:10.1007/s12275-012-1437-6.
- Schwartz VU, Hoffmann K, Nyilasi I, Papp T, Vágvolgyi C, et al. (2012) *Lichtheimia* species exhibit differences in virulence potential. *PLoS One* 7: e40908. doi:10.1371/journal.pone.0040908.
- Wang L, Chen W, Feng Y, Ren Y, Gu Z, et al. (2011) Genome characterization of the oleaginous fungus *Mortierella alpina*. *PLoS One* 6: e28319. doi:10.1371/journal.pone.0028319.
- Wöstemeier J, Kreibich A (2002) Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Curr Genet* 41: 189–198. doi:10.1007/s00294-002-0306-y.
- Ma L-J, Ibrahim AS, Skory C, Grabherr MG, Burger G, et al. (2009) Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet* 5: e1000549. doi:10.1371/journal.pgen.1000549.
- Buchon N, Vaury C (2006) RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredit (Edinb)* 96: 195–202. doi:10.1038/sj.hdy.6800789.
- James TY, Litvinseva AP, Vilgalys R, Morgan JAT, Taylor JW, et al. (2009) Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. *PLoS Pathog* 5: e1000458. doi:10.1371/journal.ppat.1000458.
- Xu X, Pan S, Cheng S, Zhang B, Mu D, et al. (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195. doi:10.1038/nature10158.
- Deutscher MP (1984) Processing of tRNA in prokaryotes and eukaryotes. *Crit Rev Biochem Mol Biol* 17: 45–71.
- Kachouri R, Strubinskis V, Zhu Y, Ramos KS, Westhof E, et al. (2005) A surprisingly large RNase P RNA in *Candida glabrata*. *RNA* 11: 1064–1072. doi:10.1261/rna.2130705.
- Schmitt ME, Clayton D a (1993) Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 13: 7935–7941.
- Walter P, Ibrahim I, Blobel G (1981) Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J Cell Biol* 91: 545–550.
- Dieci G, Fiorino G, Castelnovo M, Teichmann M, Pagano A (2007) The expanding RNA polymerase III transcriptome. *Trends Genet* 23: 614–622. doi:10.1016/j.tig.2007.09.001.

37. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24: 637–644. doi:10.1093/bioinformatics/btn013.
38. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676. doi:10.1093/bioinformatics/bt1610.
39. Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, et al. (2011) PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res* 39: D556–60. doi:10.1093/nar/gkq1109.
40. Gabaldón T (2008) Large-scale assignment of orthology: back to phylogenetics? *Genome Biol* 9: 235. doi:10.1186/gb-2008-9-10-235.
41. Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 6: 361–375. doi:10.1038/nrg1603.
42. Wehe A, Bansal MS, Burleigh JG, Eulenstein O (2008) DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics* 24: 1540–1541. doi:10.1093/bioinformatics/btn230.
43. Böcker S, Jahn K, Mixtacki J, Stoye J (2009) Computation of median gene clusters. *J Comput Biol* 16: 1085–1099. doi:10.1089/cmb.2009.0098.
44. Jahn K. (2011) Efficient computation of approximate gene clusters based on reference occurrences. *J Comput Biol* 18: 1255–1274. doi:10.2144/000113833.
45. Marcet-Houben M, Marceddu G, Gabaldón T (2009) Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evol Biol* 9: 295. doi:10.1186/1471-2148-9-295.
46. Huerta-Cepas J, Dopazo J, Gabaldón T (2010) ETE: a python Environment for Tree Exploration. *BMC Bioinformatics* 11: 24. doi:10.1186/1471-2105-11-24.
47. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–301. doi:10.1093/nar/gkr1065.
48. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–37. doi:10.1093/nar/gkr367.
49. Paoletti M, Saupé SJ, Clave C (2007) Genesis of a Fungal Non-Self Recognition Repertoire. *PLoS One*: e283. doi:10.1371/Journal.pone.00283.
50. Alastruey-Izquierdo A, Cuesta I, Walther G, Cuenca-Estrella M, Rodríguez-Tudela JL (2010) Antifungal susceptibility profile of human-pathogenic species of *Lichtheimia*. *Antimicrob Agents Chemother* 54: 3058–3060. doi:10.1128/AAC.01270-09.
51. Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol* 4: 10. doi:10.1186/1471-2229-4-10.
52. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol* 148: 993–1003. doi:10.1104/pp.108.122457.
53. Hirschman JE, Balakrishnan R, Christie KR, Costanzo MC, Dwight SS, et al. (2006) Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res* 34: D442–5. doi:10.1093/nar/gkj117.
54. Kim E, Magen A, Ast G (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* 35: 125–131. doi:10.1093/nar/gkl924.
55. Marshall AN, Montalegre MC, Jiménez-López C, Lorenz MC, van Hoof A (2013) Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. *PLoS Genet* 9: e1003376. doi:10.1371/journal.pgen.1003376.
56. Schrettl M, Bignell E, Kragl C, Jochl C, Rogers T, et al. (2004) Siderophore biosynthesis but not reductive iron assimilation is essential for *Aspergillus fumigatus* virulence. *J Exp Med* 200: 1213–1219. doi:10.1084/jem.20041242.
57. Ramanan N (2000) A High-Affinity iron permease essential for *Candida albicans* virulence. *Science* (80-): 288: 1062–1064. doi:10.1126/science.288.5468.1062.
58. Howard DH (1999) Acquisition, transport, and storage of iron by pathogenic fungi. *Clin Microbiol Rev* 12: 394–404.
59. Fu Y, Lee H, Collins M, Tsai H-F, Spellberg B, et al. (2004) Cloning and functional characterization of the *Rhizopus oryzae* high affinity iron permease (rFTR1) gene. *FEMS Microbiol Lett* 235: 169–176. doi:10.1016/j.femsle.2004.04.031.
60. Ibrahim AS, Gebremariam T, Lin L, Luo G, Husseiny MI, et al. (2010) The high affinity iron permease is a key virulence factor required for *Rhizopus oryzae* pathogenesis. *Mol Microbiol* 77: 587–604. doi:10.1111/j.1365-2958.2010.07234.x.
61. Symeonidis S (2009) The role of iron and iron chelators in zygomycosis. *Clin Microbiol Infect* 15 Suppl 5: 26–32. doi:10.1111/j.1469-0691.2009.02976.x.
62. Ibrahim AS, Spellberg B, Walsh TJ, Kontoyiannis DP (2012) Pathogenesis of Mucormycosis. *Clin Infect Dis* 54: 1–7. doi:10.1093/cid/cir865.
63. Ziegler L, Terzulli A, Gaur R, McCarthy R, Kosman DJ (2011) Functional characterization of the ferroxidase, permease high-affinity iron transport complex from *Candida albicans*. *Mol Microbiol* 81: 473–485. doi:10.1111/j.1365-2958.2011.07704.x.
64. Thielen a, Winkelmann G (1992) Rhizoferrin: a complexone type siderophore of the Mucorales and entomophthorales (Zygomycetes). *FEMS Microbiol Lett* 73: 37–41.
65. Drechsel H, Tschierske M, Thielen a, Jung G (1995) The carboxylate type siderophore rhizoferrin and its analogs produced by directed fermentation. *J Ind Microbiol* 14: 105–112.
66. Boelaert JR, de Locht M, Van Cutsem J, Kerrels V, Cantinieux B, et al. (1993) Mucormycosis during deferoxamine therapy is a siderophore-mediated infection: *In vitro* and *in vivo* animal studies. *J Clin Invest* 91: 1979–1986. doi:10.1172/JCI116419.
67. Ibrahim AS, Gebremariam T, Fu Y, Lin L, Husseiny MI, et al. (2007) The iron chelator deferasirox protects mice from mucormycosis through iron starvation. *J Clin Invest* 117: 2649–2657. doi:10.1172/JCI32338.
68. Schmitt I, Partida-Martinez LP, Winkler R, Voigt K, Einax E, et al. (2008) Evolution of host resistance in a toxin-producing bacterial-fungal alliance. *ISME J* 2: 632–641. doi:10.1038/ismej.2008.19.
69. Partida-Martinez LP, de Loos CF, Ishida K, Ishida M, Roth M, et al. (2007) Rhizonin, the first mycotoxin isolated from the zygomycota, is not a fungal metabolite but is produced by bacterial endosymbionts. *Appl Environ Microbiol* 73: 793–797. doi:10.1128/AEM.01784-06.
70. Larcher G, Dias M, Razafimandimby B, Bomal D, Bouchara J-P (2013) Siderophore production by pathogenic Mucorales and uptake of deferoxamine B. *Mycopathologia*. doi:10.1007/s11046-013-9693-5.
71. Carrano CJ, Böhnke R, Matzanke BF (1996) Fungal ferritins: the ferritin from mycelia of *Absidia spinosa* is a bacterioferritin. *FEBS Lett* 390: 261–264.
72. Capella-Gutierrez S, Marcet-Houben M, Gabaldón T (2012) Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biol* 10: 47. doi:10.1186/1741-7007-10-47.
73. Haas H, Eisendle M, Turgeon BG (2008) Siderophores in fungal physiology and virulence. *Annu Rev Phytopathol* 46: 149–187. doi:10.1146/annurev-phyto.45.062806.094338.
74. Silva MG, Schrank A, Bailão EFLC, Bailão AM, Borges CL, et al. (2011) The homeostasis of iron, copper, and zinc in paracoccidioides brasiliensis, *Cryptococcus neoformans* var. *grubii*, and *Cryptococcus gattii*: a comparative analysis. *Front Microbiol* 2: 49. doi:10.3389/fmicb.2011.00049.
75. Haas H (2012) Iron - a key nexus in the virulence of *Aspergillus fumigatus*. *Front Microbiol* 3: 28. doi:10.3389/fmicb.2012.00028.
76. Hwang LH, Seth E, Gilmore S a, Sil A (2012) SRE1 regulates iron-dependent and -independent pathways in the fungal pathogen *Histoplasma capsulatum*. *Eukaryot Cell* 11: 16–25. doi:10.1128/EC.05274-11.
77. Schoen C, Reichard U, Monod M, Kratzin HD, Rüchel R (2002) Molecular cloning of an extracellular aspartic proteinase from *Rhizopus microsporus* and evidence for its expression during infection. *Med Mycol* 40: 61–71.
78. Spreer A, Rüchel R, Reichard U (2006) Characterization of an extracellular subtilisin protease of *Rhizopus microsporus* and evidence for its expression during invasive rhino-orbital mycosis. *Med Mycol* 44: 723–731. doi:10.1080/13693780600936399.
79. Schaller M, Borelli C, Korting HC, Hube B (2005) Hydrolytic enzymes as virulence factors of *Candida albicans*. *Mycoses* 48: 365–377. doi:10.1111/j.1439-0507.2005.01165.x.
80. Park S-Y, Choi J, Lim S-E, Lee G-W, Park J, et al. (2013) Global expression profiling of transcription factor genes provides new insights into pathogenicity and stress responses in the rice blast fungus. *PLoS Pathog* 9: e1003350. doi:10.1371/journal.ppat.1003350.
81. Shelest E (2008) Transcription factors in fungi. *FEMS Microbiol Lett* 286: 145–151. doi:10.1111/j.1574-6968.2008.01293.x.
82. Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci* 5: 199–206.
83. Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: a perspective from the fungal kingdom. *Microbiol Mol Biol Rev* 74: 298–340. doi:10.1128/MMBR.00005-10.
84. Messungy F, Dubois E (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* 316: 1–21. doi:10.1016/S0378-1119(03)00747-9.
85. Goodrich JA, Tjian R (2011) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* 11: 549–558. doi:10.1038/nrg2847.
86. Leach MD, Cowen LE (2013) Surviving the heat of the moment: a fungal pathogens perspective. *PLoS Pathog* 9: e1003163. doi:10.1371/journal.ppat.1003163.
87. Haselwandter K, Ebner MR (1994) Microorganisms surviving for 5300 years. *FEMS Microbiol Lett* 116: 189–193.
88. Chevreaux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*. pp. 45–56.
89. Sommer DD, Delcher AL, Salzberg SL, Pop M (2007) Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 8: 64. doi:10.1186/1471-2105-8-64.
90. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5: R12. doi:10.1186/gb-2004-5-2-r12.
91. Jurka J, Kapitonov V V, Pavlicek a, Klonowski P, Kohany O, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110: 462–467. doi:10.1159/000084979.

92. Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. *BMC Bioinformatics* 7: 474. doi:10.1186/1471-2105-7-474.
93. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955–964.
94. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, et al. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35: 3100–3108. doi:10.1093/nar/gkm160.
95. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, et al. (2011) Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Res* 39: D141–5. doi:10.1093/nar/gkq1129.
96. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. doi:10.1016/S0022-2836(05)80360-2.
97. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25: 1335–1337. doi:10.1093/bioinformatics/btp157.
98. Gautheret D, Major F, Cedergren R (1990) Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *CABIOS* 6: 325–331.
99. Timothy TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*. pp. 28–36.
100. Griebel T, Brinkmeyer M, Böcker S (2008) EPoS: a modular software framework for phylogenetic analysis. *Bioinformatics* 24: 2399–2400. doi:10.1093/bioinformatics/btn364.
101. Kong Y (2011) Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* 98: 152–153. doi:10.1016/j.ygeno.2011.05.009.
102. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14: R36. doi:10.1186/gb-2013-14-4-r36.
103. Wöstemeyer J (1985) Strain-dependent variation in ribosomal DNA arrangement in *Absidia glauca*. *Eur J Biochem* 146: 443–448.
104. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140. doi:10.1093/bioinformatics/btp616.
105. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113. doi:10.1186/1471-2105-5-113.
106. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33: 511–518. doi:10.1093/nar/gki198.
107. Landan G, Graur D (2007) Heads or tails: a simple reliability check for multiple sequence alignments. *Mol Biol Evol* 24: 1380–1383. doi:10.1093/molbev/msm060.
108. Wallace IM, O’Sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res* 34: 1692–1699. doi:10.1093/nar/gkl091.
109. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973. doi:10.1093/bioinformatics/btp348.
110. Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* 14: 685–695.
111. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321. doi:10.1093/sysbio/syq010.
112. Akaike H (1973) Information theory and extension of the maximum likelihood principle. *Proceedings of the 2nd international symposium on information theory*. pp. 267–281.
113. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25: 1307–1320. doi:10.1093/molbev/msn067.
114. Schmidt T, Stoye J (2007) Gecko and GhostFam. *Methods in Molecular Biology*. Vol. 396. pp. 165–182. doi:10.1007/978-1-59745-515-2_12.
115. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, et al. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438: 1151–1156. doi:10.1038/nature04332.
116. Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, et al. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438: 1105–1115. doi:10.1038/nature04341.
117. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414: 450–453. doi:10.1038/35106579.
118. Wood V, Gwilliam R, Rajandream MA, Linc M, Linc R, et al. (2002) Genome sequence of *Schizosaccharomyces pombe*. *Nature* 415: 871–880.
119. Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, et al. (2009) Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathog* 5: e1000466. doi:10.1371/journal.ppat.1000466.
120. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, et al. (2012) Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth: 2478–2488. doi:10.1101/gr.142802.112.2478.
121. Duplessis S, Cuomo CA, Lin Y, Aerts A, Tisserant E, et al. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. doi: 10.1073/pnas.1019315108
122. Stajich JE, Wilke SK, Ahrén D, Au CH, Birren BW, et al. (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc Natl Acad Sci U S A* 107: 11889–11894. doi:10.1073/pnas.1003391107.
123. Martínez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, et al. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* 22: 695–700. doi:10.1038/nbt967.
124. Martin F, Aerts a, Ahrén D, Brun a, Danchin EGJ, et al. (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452: 88–92. doi:10.1038/nature06556.
125. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86–94. doi:10.1126/science.1139158.
126. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, et al. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451: 783–788. doi:10.1038/nature06617.
127. Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, et al. (2011) The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333: 762–765. doi:10.1126/science.1205411.
128. Joneson S, Stajich JE, Shiu S-H, Rosenblum EB (2011) Genomic transition to pathogenicity in chytrid fungi. *PLoS Pathog* 7: e1002338. doi:10.1371/journal.ppat.1002338.
129. Galagan JE, Calvo SE, Borkovich K a, Selker EU, Read ND, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859–868. doi:10.1038/nature01554.
130. Stamatakis a, Ludwig T, Meier H (2005) RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456–463. doi:10.1093/bioinformatics/bti191.