Report

Current Biology

Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication

Highlights

- We sequenced the genomes of fungi *Mucor circinelloides* and *Phycomyces blakesleeanus*
- Extensive genomic duplications increased the number of genes for signal transduction
- After duplication, genes specialized their transcriptional regulation by light
- Genome duplications provided new components to improve sensory perception in fungi

Authors

Luis M. Corrochano, Alan Kuo, Marina Marcet-Houben, ..., Alfredo Herrera-Estrella, Toni Gabaldón, Igor V. Grigoriev

Correspondence

corrochano@us.es

In Brief

Fungi use light and other signals to regulate development, but the mechanisms and origin are largely unknown. Corrochano et al. have found that a group of fungi has acquired new components for signal transduction pathways after extensive genomic duplication. The expanded gene repertoire allows an enhanced perception of environmental signals.

Accession Numbers

AMYC00000000
AMYB00000000
KR809878
KR809877
GSE64369
GSE58264

Corrochano et al., 2016, Current Biology *26*, 1–8
June 20, 2016 © 2016 Elsevier Ltd.
http://dx.doi.org/10.1016/j.cub.2016.04.038

CellPress

Please cite this article in press as: Corrochano et al., Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.04.038

Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication

Luis M. Corrochano,[1,*] Alan Kuo,[2] Marina Marcet-Houben,[3,4] Silvia Polaino,[5] Asaf Salamov,[2] José M. Villalobos-Escobedo,[6] Jane Grimwood,[2,7] M. Isabel Álvarez,[8] Javier Avalos,[1] Diane Bauer,[2] Ernesto P. Benito,[8,9] Isabelle Benoit,[10] Gertraud Burger,[11] Lola P. Camino,[1] David Cánovas,[1] Enrique Cerdá-Olmedo,[1] Jan-Fang Cheng,[2] Angel Domínguez,[8] Marek Eliáš,[12] Arturo P. Eslava,[8] Fabian Glaser,[13] Gabriel Gutiérrez,[1] Joseph Heitman,[14] Bernard Henrissat,[15,16] Enrique A. Iturriaga,[8] B. Franz Lang,[11] José L. Lavín,[17] Soo Chan Lee,[14] Wenjun Li,[14] Erika Lindquist,[2] Sergio López-García,[18] Eva M. Luque,[1] Ana T. Marcos,[1] Joel Martin,[2] Kevin McCluskey,[19]

(Author list continued on next page)

[1]Department of Genetics, University of Seville, Avenida Reina Mercedes s/n, 41012 Seville, Spain
[2]US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
[3]Centre for Genomic Regulation (CRG), Doctor Aiguader 88, 08003 Barcelona, Spain
[4]Universitat Pompeu Fabra (UPF), Doctor Aiguader 88, 08003 Barcelona, Spain
[5]School of Biological Sciences, University of Missouri–Kansas City, 5007 Rockhill Road, Kansas City, MO 64110, USA
[6]Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), Centro de Investigación y Estudios Avanzados, Kilómetro 9.6 Libramiento Norte, Carretera Irapuato-León, 36821 Irapuato, Guanajuato, México
[7]HudsonAlpha Institute of Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806, USA
[8]Departamento de Microbiología y Genética, Universidad de Salamanca, Plaza de los doctores de la Reina s/n, 37007 Salamanca, Spain
[9]Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), Universidad de Salamanca, Río Duero 12, 37185 Salamanca, Spain
[10]CBS-KNAW Fungal Biodiversity Centre and Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
[11]Universite de Montreal, Pavillon Roger-Gaudry, Biochimie, CP 6128, Succursale Centre-Ville, Montreal QC, H3C 3J7, Canada
[12]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic
[13]Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion - Israel Institute of Technology, Haifa 3200003, Israel
[14]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Research Drive, Durham, NC 27710, USA
[15]Centre National de la Recherche Scientifique (CNRS), UMR7257, Université Aix-Marseille, 163 Avenue de Luminy, 13288 Marseille, France
[16]Department of Biological Sciences, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia

(Affiliations continued on next page)

SUMMARY

Plants and fungi use light and other signals to regulate development, growth, and metabolism. The fruiting bodies of the fungus *Phycomyces blakesleeanus* are single cells that react to environmental cues, including light, but the mechanisms are largely unknown [1]. The related fungus *Mucor circinelloides* is an opportunistic human pathogen that changes its mode of growth upon receipt of signals from the environment to facilitate pathogenesis [2]. Understanding how these organisms respond to environmental cues should provide insights into the mechanisms of sensory perception and signal transduction by a single eukaryotic cell, and their role in pathogenesis. We sequenced the genomes of *P. blakesleeanus* and *M. circinelloides* and show that they have been shaped by an extensive genome duplication or, most likely, a whole-genome duplication (WGD), which is rarely observed in fungi [3–6]. We show that the genome duplication has expanded gene families, including those involved in signal transduction, and that duplicated genes have specialized, as evidenced by differences in their regulation by light. The transcriptional response to light varies with the developmental stage and is still observed in a photoreceptor mutant of *P. blakesleeanus*. A phototropic mutant of *P. blakesleeanus* with a heterozygous mutation in the photoreceptor gene *madA* demonstrates that photosensor dosage is important for the magnitude of signal transduction. We conclude that the genome duplication provided the means to improve signal transduction for enhanced perception of environmental signals. Our results will help to understand the role of genome dynamics in the evolution of sensory perception in eukaryotes.

RESULTS AND DISCUSSION

Genome Duplications in the Evolution of the Mucoromycotina Fungi

Gene duplication has expanded the number of genes for photoreception in *P. blakesleeanus* and *M. circinelloides* (Figures 1A and 1B) [7–9], and we hypothesize that gene duplications and specialization may have provided new proteins to expand their sensory repertoire. We thus sequenced the 53.9-Mb *P. blakesleeanus* and the 36.6-Mb *M. circinelloides* genomes,

Humberto R. Medina,[1] Alejandro Miralles-Durán,[1] Atsushi Miyazaki,[20] Elisa Muñoz-Torres,[21] José A. Oguiza,[22] Robin A. Ohm,[2] María Olmedo,[1] Margarita Orejas,[23] Lucila Ortiz-Castellanos,[24] Antonio G. Pisabarro,[22] Julio Rodríguez-Romero,[1] José Ruiz-Herrera,[24] Rosa Ruiz-Vázquez,[18] Catalina Sanz,[8] Wendy Schackwitz,[2] Mahdi Shahriari,[8] Ekaterina Shelest,[25] Fátima Silva-Franco,[18] Darren Soanes,[26] Khajamohiddin Syed,[27] Víctor G. Tagua,[1] Nicholas J. Talbot,[26] Michael R. Thon,[8,9] Hope Tice,[2] Ronald P. de Vries,[10] Ad Wiebenga,[10] Jagjit S. Yadav,[27] Edward L. Braun,[28] Scott E. Baker,[29] Victoriano Garre,[18] Jeremy Schmutz,[2,7] Benjamin A. Horwitz,[30] Santiago Torres-Martínez,[18] Alexander Idnurm,[5,31] Alfredo Herrera-Estrella,[6] Toni Gabaldón,[3,4,32] and Igor V. Grigoriev[2]
[17]Genome Analysis Platform, CIC bioGUNE, Bizkaia Technology Park, 48160 Derio, Bizkaia, Spain
[18]Departamento de Genética y Microbiología, Universidad de Murcia, 30071 Murcia, Spain
[19]Department of Plant Pathology, Kansas State University, 4024 Throckmorton Plant Sciences Center, Manhattan, KS 66506, USA
[20]Department of Biological Sciences, Faculty of Science and Engineering, Ishinomaki Senshu University, Ishinomaki 986-8580, Japan
[21]Departamento de Biología Celular y Patología, Facultad de Medicina, Universidad de Salamanca, Avenida Campus Miguel de Unamuno, 37007 Salamanca, Spain
[22]Department of Agrarian Production, Public University of Navarre, 31006 Pamplona, Spain
[23]Instituto de Agroquímica y Tecnología de Alimentos, Consejo Superior de Investigaciones Científicas (IATA-CSIC), Avenida Catedrático Agustín Escardino 7, 46980 Paterna, Valencia, Spain
[24]Departamento de Ingeniería Genética, Unidad Irapuato, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Kilómetro 9.6 Libramiento Norte, Carretera Irapuato-León, 36821 Irapuato, Guanajuato, Mexico
[25]Leibniz Institute for Natural Product Research and Infection Biology (Hans Knoell Institute), Beutenbergstrasse 11a, 07745 Jena, Germany
[26]Biosciences, College of Life and Environmental Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, UK
[27]Department of Environmental Health, University of Cincinnati College of Medicine, 160 Panzeca Way, Cincinnati, OH 45267-0056, USA
[28]Department of Biology, University of Florida, P.O. Box 118525, Gainesville, FL 32611-8525, USA
[29]Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, USA
[30]Department of Biology, Technion - Israel Institute of Technology, Haifa 32000, Israel
[31]School of BioSciences, University of Melbourne, Melbourne, VIC 3010, Australia
[32]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain
*Correspondence: corrochano@us.es
http://dx.doi.org/10.1016/j.cub.2016.04.038

and their respective mtDNAs (Supplemental Experimental Procedures; Tables S1A–S1D; Figure S1). The increase in genome size in *P. blakesleeanus* is, in part, due to repetitive DNA, including transcribed transposable elements (Tables S1E–S1G). 16,528 (*P. blakesleeanus*) and 11,719 (*M. circinelloides*) protein-coding genes were annotated and compared to proteins from other fungi (Figure 1C). Comparison of the two genomes with that of *Rhizopus delemar* and other fungi suggests that a whole-genome duplication (WGD) occurred early in the Mucoromycotina lineage. The fungal kingdom contains the subkingdom Dikarya and a number of early divergent lineages including the Mucoromycotina with *P. blakesleeanus*, *M. circinelloides*, and *R. delemar* [10]. Two WGDs have been reported in fungi: in the Saccharomycotina, a lineage of the Dikarya [4–6], and in *R. delemar* [3].

Ancient WGDs are difficult to detect because gene loss and rearrangements result in the absence of regions of synteny. However, genomes from Mucoromycotina species have more members per gene family than genomes from Dikarya fungi (2.9–3.6 versus 1.6–2.2) and a large fraction of gene families with more members than average (50%–68% versus 6.9%–22% for each Dikarya species) (Supplemental Experimental Procedures; Table S2A). Moreover, Mucoromycotina genomes have more duplicated regions than other fungal genomes, with four to 13 genes on average (Tables S2B and S2C). We confirmed the presence of duplicated regions after WGDs in the genomes of *S. cerevisiae* and *R. delemar* as expected, and also large amounts of duplicated DNA in several Dikarya fungi (Tables S2B and S2C). The *Puccinia graminis* and *Laccaria bicolor* genomes have expanded lineage-specific gene families proposed to be involved in pathogenesis and symbiosis [11, 12]. These duplicated regions contain large fractions of lineage-specific genes (Table S2C) supporting the proposal that they arose after species-specific segmental duplications. Additional WGD signatures can be observed in families of three genes from genome pairs. These types of duplicates are more frequent in the genomes of Mucoromycotina species than in non-Mucoromycotina species, suggesting that the former harbors traces of past WGDs (Table S2D).

To gain further insight into past genome expansions in the Mucoromycotina, we reconstructed the complete collection of evolutionary histories (i.e., the phylome) for genes within Mucoromycotina fungi with 13 other fungal genomes. The gene trees were analyzed to detect and date duplication events [13] (Figure 1C). This method has been used to characterize the WGD that took place in the *S. cerevisiae* lineage [14]. In addition to the WGD described in *R. delemar* (0.43 duplications per gene), we detected a larger duplication peak (0.70–0.96 duplications per gene) in the lineage preceding the Mucoromycotina species consistent with a WGD preceding the diversification of this lineage. This early WGD explains gene duplications in the oxidative phosphorylation complex in Mucorales [15] and segmental duplications in *Lichtheimia corymbifera* where previously a species-specific WGD had been rejected [16].

The best explanation for our observations is a WGD predating the diversification of the Mucorales followed by a WGD in the *R. delemar* lineage and, subsequently, rampant gene loss as observed in yeast [5]. The alternative explanation (lineage-specific gene duplications) is less parsimonious. Although segmental duplications can create paralogous regions with shared synteny, we consider it unlikely that numerous such events affecting large regions of the genome would have coincided in time or affected multiple lineages in parallel.

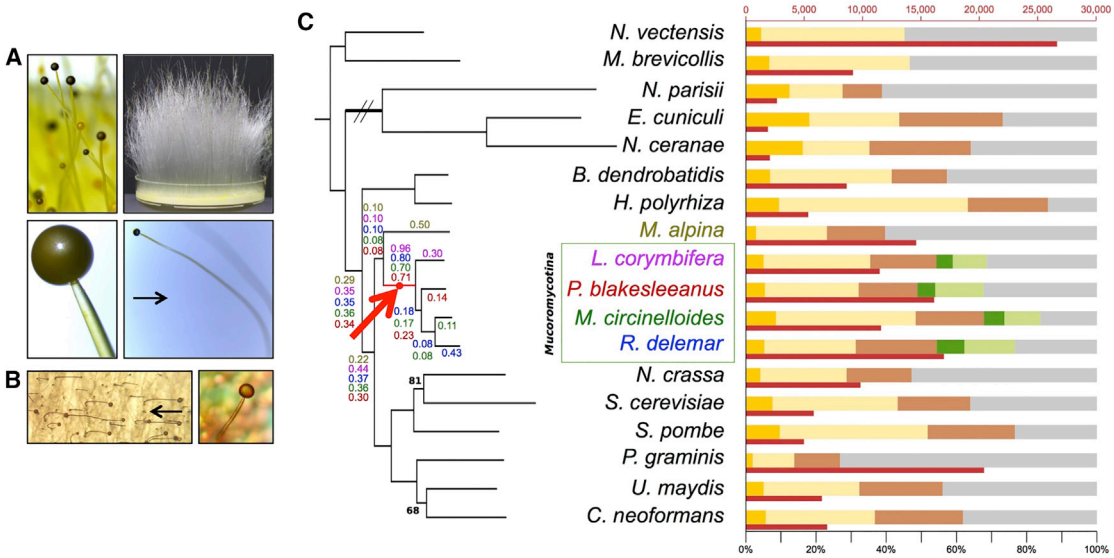

Figure 1. Sensory Perception and Genome Duplication in the Mucoromycotina

(A) The fruiting bodies, sporangiophores, of *Phycomyces blakesleeanus* grow out of the mycelium and reach several centimeters in length. The speed and direction of growth is controlled by signals from the environment including light, gravity, touch, wind, and the presence of nearby objects. The ball at the top of each fruiting body is the sporangium with spores. The direction of light is indicated by an arrow.

(B) The sporangiophores of *Mucor circinelloides* are small (about 5 mm) and show phototropism. The direction of light is indicated by an arrow.

(C) Evidence for a WGD in the Mucoromycotina. A fungal evolutionary tree with bootstrap support values lower than 95% indicated in black numbers at the branches. The average duplication per gene in each lineage is shown with a color that indicates the phylome used for the duplication density calculation. The branch where the proposed WGD took place is marked in red with a dot and an arrow. The graph (scale on the bottom) represents the percentage of genes in a given species that belong to one of the following categories: yellow, protein present in all species; light yellow, ancestral proteins that have homologs in the outgroups; brown, Fungi-specific proteins; green, Mucoromycotina-specific proteins, which appear in all four species; light green, Mucoromycotina-specific proteins; gray, species-specific proteins. The red bars (scale on top) represent the total number of proteins encoded in each genome.

See also Figure S1 and Tables S1 and S2.

Expansion of Gene Families in Mucoromycotina Fungi

Duplicated genes in the four Mucoromycotina species contained an abundance of gene ontology (GO) terms for protein kinase activities (GO 4674, GO 4672, GO 4713, GO 4707; p value at least 1×10^{-6}), fructose 2,6-bisphosphate metabolic process (GO 6003; $p = 1.3 \times 10^{-6}$), ATP binding (GO 5524; $p = 1.8 \times 10^{-46}$), and protein transport (GO 15031; $p = 1.6 \times 10^{-17}$), suggesting duplications of genes encoding signaling pathways and transport components (Supplemental Experimental Procedures; Figure 2; Data S1). This is supported by further analysis of the abundance of signal transduction genes. We have limited this analysis to the genomes of *P. blakesleeanus*, *M. circinelloides*, and *R. delemar* as they were the only Mucoromycotina genomes sequenced when we started the project.

Heterotrimeric G protein signaling is central to the life cycle and virulence of fungi [17, 18]. All the gene families encoding the subunits of heterotrimeric G proteins are expanded. The Gα subunit family contains ten, 12, and 12 genes in *P. blakesleeanus*, *M. circinelloides*, and *R. delemar*, respectively, compared to an average of three in the Ascomycota or seven in the Dikarya (Figure 2A). A single Gβ subunit gene has been found in the Ascomycota as well as in the basidiomycetes *Ustilago maydis* and *Cryptococcus neoformans* [19, 20]. In *P. blakesleeanus*, we identified five Gβ genes, in *M. circinelloides* three, and in *R. delemar* four. Similarly, the Mucoromycotina genomes have three or four genes for the Gα subunit compared to an average of one in the Dikaria (Figure 2A). Theoretically, a very large number of G protein heterotrimers could be built from the multiple subunits. An estimate of 21, 21, and ten genes for G-protein-coupled receptors (GPCR) in *P. blakesleeanus*, *M. circinelloides*, and *R. delemar*, respectively (Data S1, sheet 1), suggests moderate expansion compared to ten to 12 genes in Dikarya [21].

The number of other signal transduction genes has increased compared to the Dikarya, including genes for protein kinases, TRAFAC class GTPases, and regulators of GTPases of the Ras superfamily (Data S1, sheets 2–4). The expansion of kinase families is 3- to 4-fold, as there are 63, 70, and 82 CAMK genes in *P. blakesleeanus*, *M. circinelloides*, and *R. delemar*, respectively, compared to 22 and 21 in *Neurospora crassa* and *U. maydis*. Other families show larger expansion, e.g., 11–18 genes for casein kinase 1 in Mucoromycotina compared to two to three in Dikarya (Figure 2A; Data S1, sheet 4).

Some, but not all, families in a given category are expanded; e.g., for photoreception [22], the genes for components of the WC photoreceptor complex (WC-1 and WC-2) are duplicated, but not the cryptochrome gene; genes for casein kinase 1 are duplicated, but not those encoding the sensor histidine kinases (Figure 2A). Genes for calcium or pH sensing show non-uniform duplication: there are multiple calmodulin genes and three calcineurin catalytic subunit genes, but only a single calcineurin regulatory subunit; in the pH pathway there are three to four genes for the PacC transcription factor and two genes for PalA, yet one gene for PalB or PalC as in the ascomycete *Aspergillus nidulans* [23] (Data S1, sheet 1). Cyclin families are expanded, but there is

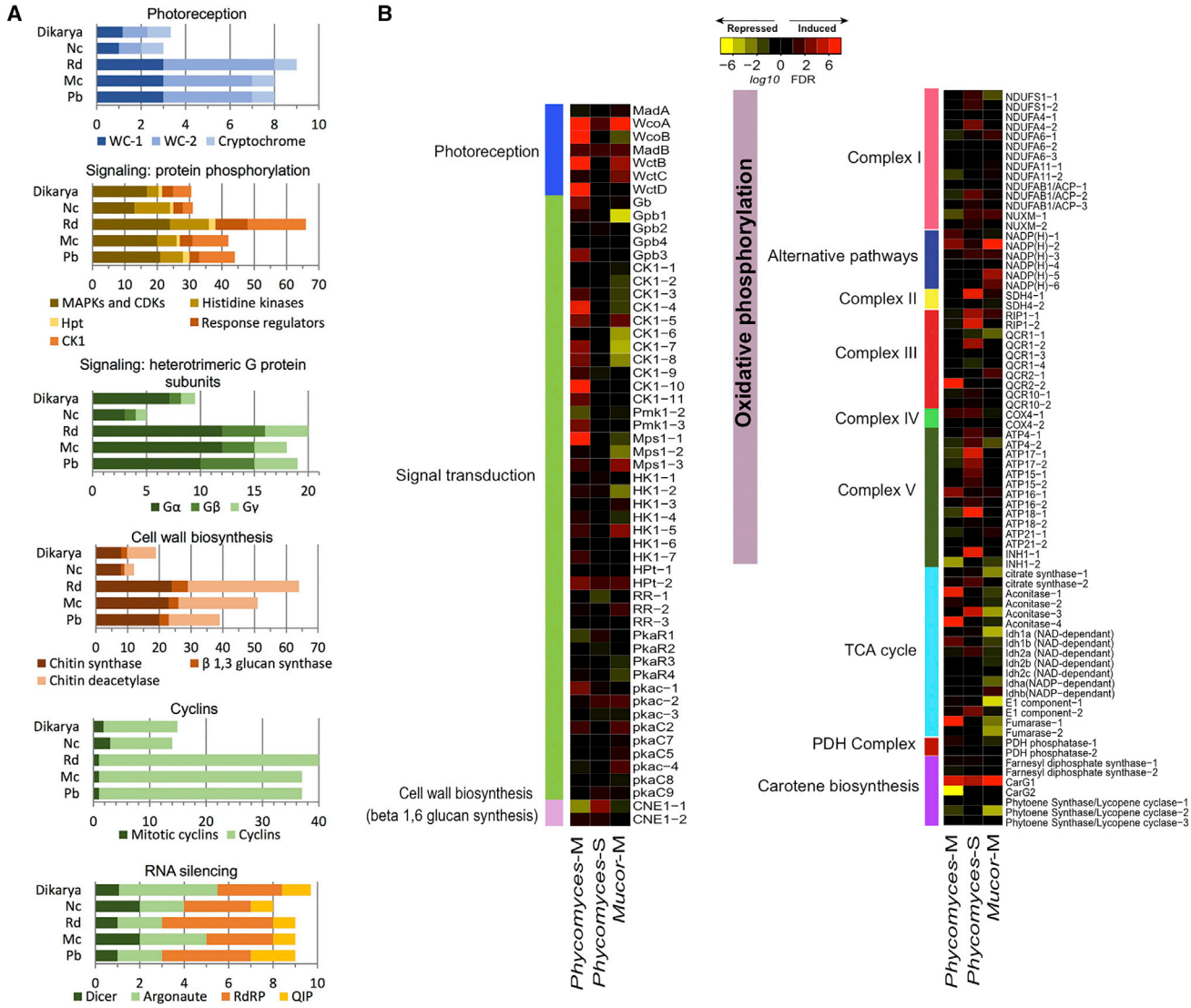

Figure 2. Gene Expansion and Transcriptional Specialization in the Mucoromycotina

(A) Gene abundance in Mucoromycotina, *Neurospora crassa*, and Dikarya fungi. The x axis indicates number of genes, and the bars, from bottom to top, indicate numbers of predicted genes for the three Mucoromycotina (Pb, *P. blakesleeanus*; Mc, *M. circinelloides*; Rd, *R. delemar*), and average number of genes for *N. crassa* (Nc), and Dikarya.

(B) Expression patterns in response to light of duplicated genes in *P. blakesleeanus* and *M. circinelloides*. Differential expression of genes was obtained for two *P. blakesleeanus* stages (mycelium and sporangiophore), and *M. circinelloides* mycelium. The three *wc-1* genes are *madA*, *wcoA*, and *wcoB*, and the four *wc-2* genes are *madB*, *wctB*, *wctC*, and *wctD*. Results are represented with the logarithm base ten of false discovery rate (FDR) (FDR <0.05; fold change >2). M, mycelium; S, sporangiophore.

See also Figure S2 and Data S1.

only a single mitotic cyclin, compared to several in the Dikarya (Figure 2A). Not all gene families have expanded. For example, the genes encoding proteins that participate in genome defense through RNAi have not duplicated (Figure 2A) [24].

The genome duplication has multiplied genes involved in cell wall biosynthesis, in particular, chitin synthases and chitin deacetylases (Figure 2A). These enzymes may have specialized to modulate the growth response of the sporangiophore after environmental stimuli. Functional specialization after gene duplication should have played a key role in *M. circinelloides* and other pathogenic Mucoromycotina fungi. Mutants of the photoreceptor gene *wc-1* of *Fusarium oxysporum* and *C. neoformans* show decreased virulence [25, 26]. *M. circinelloides wc* genes have specialized their sensory role after gene duplication [9], suggesting that some WC proteins may serve as pathogenicity factors (S.T.-M., unpublished data). Calcineurin is a virulence factor in several fungi, including *M. circinelloides* [27]. The number of calcineurin A catalytic subunit genes has increased, with three genes in *M. circinelloides* compared to one in other fungi (Data S1, sheet 1). One of them, *cnaA*, is involved in virulence [27], confirming gene specialization to facilitate pathogenesis after duplication.

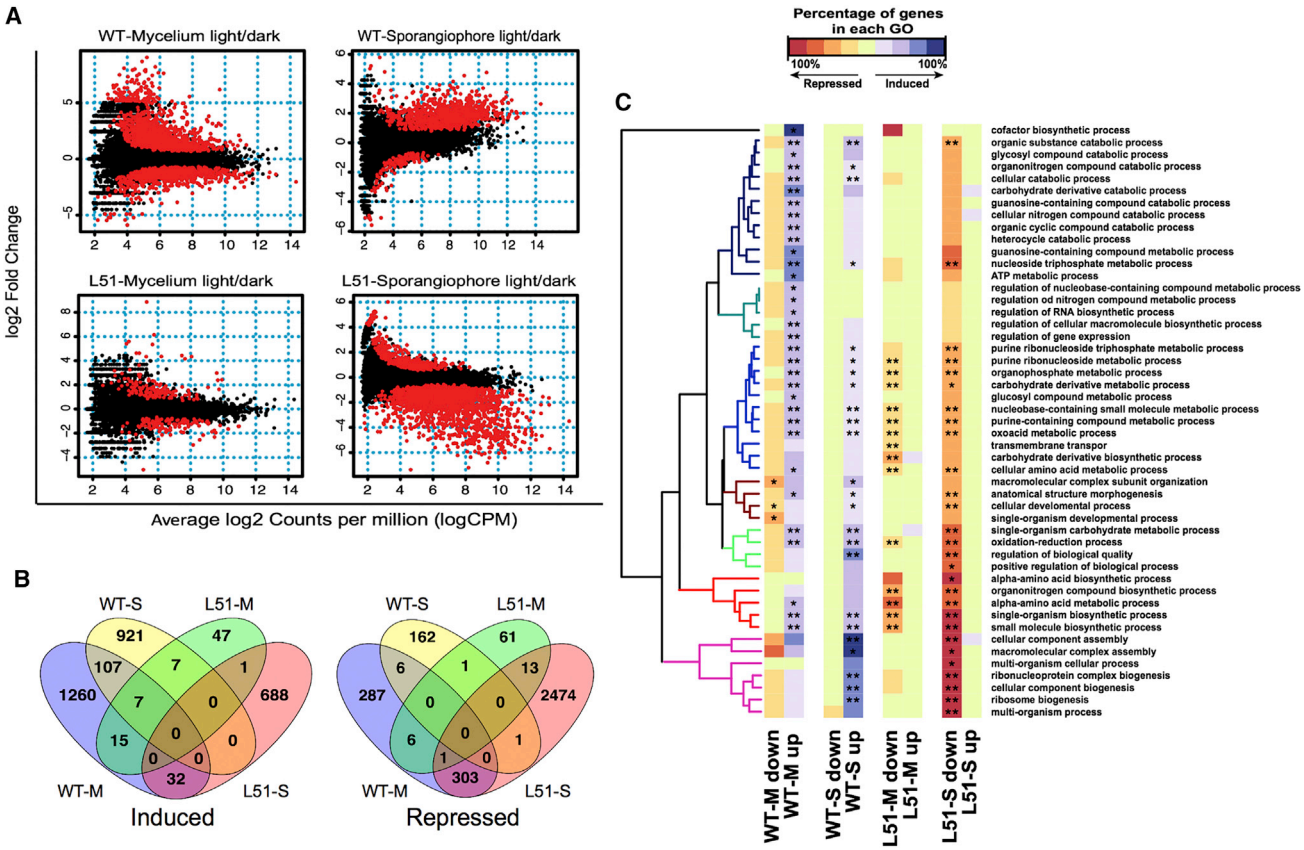

Figure 3. The Influence of Light on Gene Expression in Two Developmental Stages of *P. blakesleeanus*

(A) Differential expression (light/dark) in the mycelium or the sporangiophore of the wild-type and the *madA madB* mutant strain (L51). Differentially expressed genes with FDR ≤ 0.05 are shown in red.

(B) Overlap of genes induced and repressed in the wild-type and mutant using RNA from mycelia (WT-ML and L51-ML) or sporangiophores (WT-SL and L51-SL).

(C) Category enrichment in differentially expressed genes (*FDR <0.05; **FDR <0.01). Each vertical block contains the up- and downregulated categories. Color intensity represents the percentage of genes belonging to each category and includes only GO terms for Biological Processes. Clusters based on this percentage are displayed in different colors in the tree.

See also Data S2.

The expansion of some gene families can be accounted for by a WGD and retention of the resulting paralogs. The large number of chitin deacetylases or Gα subunits genes (Figure 2A; Data S1, sheet 1), however, cannot be explained by a WGD alone, suggesting additional segmental duplications. Having more proteins for signal transduction and cell wall biosynthesis should have helped Mucoromycotina fungi to improve environmental sensing and responses, including the perception of potential hosts for pathogenic fungi. Elucidation of the biological role of duplicated genes will require further characterization.

Duplicated Genes Differ in their Transcriptional Response to Light

To investigate whether duplicated genes have specialized, we asked whether duplicated genes from *P. blakesleeanus* and *M. circinelloides* responded differently to light. The transcriptome of cultures kept in the dark or after exposure to 30 min of blue light (2.3×10^3 J/m^2) indicates specialization (Figure 2B). Most of the genes encoding components of the photoreceptor complex (WCC) or the regulatory subunits of protein kinase A showed a similar expression pattern in *P. blakesleeanus* and *M. circinelloides* with only some genes regulated by light despite being transcriptionally active in both mycelia and sporangiophores (Figure 2B). In addition, genes encoding the photoreceptor WcoB, the Gβ subunits Gpb1 and Gpb3, and the kinases Pkac-1, Mps1-1, CK1-7, and CK1-8, showed opposite responses to light in *P. blakesleeanus* and *M. circinelloides*. Thus, duplicated genes have evolved different patterns of expression in different organisms, as well as between specific tissues in the same organism.

A Refined Transcriptional Response to Light in *P. blakesleeanus*

The expansion in the number of photoreceptors and other signal transduction proteins may have allowed fine-tuning of the response to light, for example, allowing tissue-specific transcriptional responses. We thus characterized the global transcriptional response in *P. blakesleeanus* mycelium and sporangiophores by RNA sequencing (RNA-seq) (Figure 3; Data S2). A total of 2,024 genes were responsive to light in the mycelium (1,421 induced and 603 repressed, about 12% of the protein-coding genes), compared to 1,212 genes in the sporangiophores

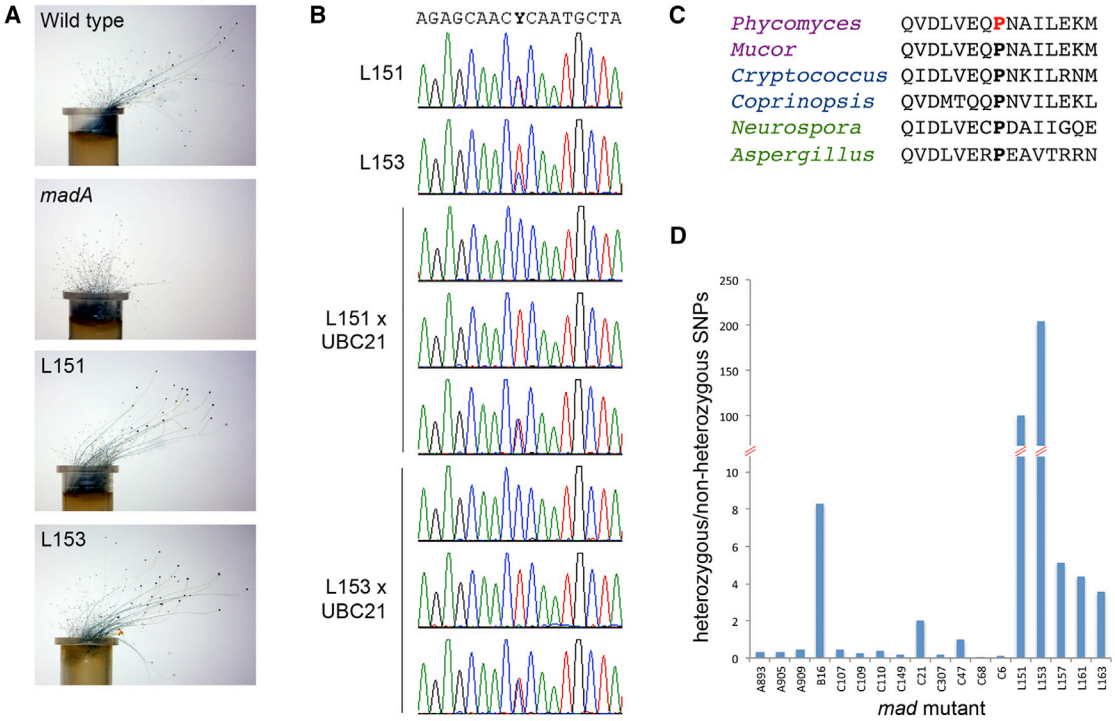

Figure 4. The *P. blakesleeanus madI* Strains Are Heterokaryons of Wild-Type and *madA* Nuclei

(A) Phenotype of wild-type, *madA*, and *madI* strains L151 and L153, with illumination from the right. The average bending response of *madI* strains is lower than in the wild-type strain with the intensity that we used.

(B) DNA sequencing chromatograms of a region of the *madA* gene from the two *madI* strains and progeny from the *madI* mutants crossed to wild-type UBC21. The progeny sequences represent the three different types observed in the strains obtained from crosses.

(C) A segment of the sequence of MadA in different fungi with the conserved proline that is mutated in *madI* strains in bold.

(D) Graph showing the ratio of heterozygous/non-heterozygous SNPs in the genomes of 19 *mad* mutants.

See also Figure S3 and Data S3.

(1,042 induced and 170 repressed). The transcriptional response to light was specific for each developmental stage because only 120 genes were light-regulated in both mycelium and sporangiophores (Figures 3A and 3B). The same analysis on a *madA madB* double mutant (strain L51) that is considered blind [28] detected only 159 light-regulated genes in the mycelium confirming the relevance of the Mad complex (Figure 3A). Surprisingly, the *madA madB* mutant showed a significant response to light in sporangiophores where 3,513 genes were regulated by light (Figure 3A). It is noteworthy that most responsive genes in the *madA madB* mutant sporangiophores were repressed by light (Figure 3A). This suggests the activity of light-dependent repressors in the absence of the Mad complex, as proposed by electrophoretic mobility shift assay (EMSA) experiments that showed the binding of proteins in the dark to a light-regulated promoter in a *madA madB* mutant [29]. Seven gene clusters were enriched in regulatory genes in light-induced mRNAs from the mycelium of the wild-type, and in ribosome biogenesis genes in the sporangiophores (Figure 3C). *P. blakesleeanus* and *M. circinelloides* have 879 and 650 genes encoding transcription factors (TFs), respectively (about 5% of the protein-coding genes), with an abundance of C_2H_2 Zn finger TFs (Figure S2). Light regulates 9% of the *P. blakesleeanus* transcription factor genes (Data S1, sheets 5–7; Data S2, sheets 5 and 6), and we propose that the stage-specific transcriptional response to light relies on the expanded set of photoreceptors and light-dependent transcriptional regulators. Specialization of genes for signal transduction following WGD has been observed in vertebrate vision [30]. Strikingly, in both vertebrates and fungi the expansion of signal transduction genes after WGD has resulted in more elaborate sensory perception.

Reduced Sensitivity to Light in Strains with Wild-Type and *madA* Mutant Nuclei

The photoresponse in the *P. blakesleeanus madA madB* mutant suggested the action of additional photoreceptors. Light perception in *madI* mutants is reduced 10- to 1,000-fold, halfway between wild-type and *madA* mutants, making MadI a candidate for a photoreceptor [31, 32] (Figure 4A). To identify *madI*, we crossed two *madI* strains with a wild-type strain and characterized the phototropism and molecular markers in the progeny [33]. Weak linkage was found for the *madI* mutation and three scaffolds, including the one carrying *madA* (Data S3). We therefore sequenced the genomes of two *madI* mutants (L151 and L153), along with another 17 *mad* mutant strains, and the sequences were scanned across the three scaffolds. We found that the two *madI* strains had an identical and unique mutation in *madA* [8] changing a conserved proline to leucine (Figures 4B and 4C). However, the *madI* strains also contained the wild-type allele, indicating that they were heterozygous for this

gene. To confirm these observations, we sequenced *madA* in 63 *madI* × wild-type progeny: eight only had the mutation in the *madA* gene, 35 were wild-type, and 20 were heterozygotes. The two *madI* mutants had a high number of heterozygous sites across their genomes compared to other strains (Figure 4D). Analysis of all scaffolds in the L151 and L153 genomes showed that the heterozygous SNPs are distributed throughout all chromosomes (Figure S3), suggesting that the two *madI* strains are heterokaryons or diploids, rather than being aneuploid or carrying a segmental duplication. The observation that two of the *madI* strains are heterozygous wild-type/*madA* mutants shows that sensitivity to light is related to the dosage of the MadA photoreceptor.

Our characterization of the genomes of *P. blakesleeanus* and *M. circinelloides*, our comparative fungal genome analysis, and our gene function studies provide new insight into the occurrence and consequences of genome duplications in the evolution of fungi. Expansion and specialization of genes for signal transduction and cell-wall biosynthesis following genome duplication in the Mucoromycotina provided new proteins that have enabled these fungi to refine the way they perceive signals from the environment to regulate their growth and development. Our results provide new genomic tools to unravel the molecular mechanisms of sensory perception in early diverging fungi that will help to understand the evolution of sensory perception in eukaryotes.

ACCESSION NUMBERS

Genome assemblies and annotations have been deposited in DDBJ/EMBL/GenBank: AMYC00000000 (*P. blakesleeanus*) and AMYB00000000 (*M. circinelloides*) and are accessible via the JGI Mycocosm portal (http://jgi.doe.gov/fungi). The mtDNAs have been deposited in DDBJ/EMBL/GenBank: KR809878 (*P. blakesleeanus*) and KR809877 (*M. circinelloides*). Phylomes can be accessed at phylomeDB (http://phylomedb.org) with phylomeIDs 252, 253, 254, 255, and 256. Gene expression data have been deposited in the GEO database: GSE64369 (*P. blakesleeanus*) and GSE58264 (*M. circinelloides*).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, two tables, and three datasets and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.04.038.

AUTHOR CONTRIBUTIONS

A.K., M.M.-H., S.P., A.S., J.M.V.-E., and J.G. contributed equally to this work. L.M.C. coordinated the project. E.L.B., S.B., S.T.-M. and L.M.C. conceived the project. I.V.G. directed all genome-sequencing and analysis efforts at JGI. J.R.-R., V.G.T., A.M.-D., C.S., M.S., E.M.L., F.S.-F., and V.G. provided DNA and RNA samples. A.K., A.S., E.L., J.-F.C., W.S., J.M., J.S., J.G., H.T., and D.B. sequenced, assembled, and annotated the genomes. M.M.-H., A.S., T.G., and I.V.G. did genome comparisons and phylogenomics analysis. M.I.A., J.A., E.P.B., I.B., G.B., L.P.C., D.C., E.C.O., L.M.C., A.D., M.E., A.P.E., V.G., F.G., G.G., J.H., B.H., B.A.H., A.I., E.A.I., B.F.L., J.L.L., S.C.L., W.L., A.T.M., H.R.M., A.M., E.M.-T., J.A.O., R.A.O., M. Olmedo, M. Orejas, L.O.-C., A.G.P., J.R-R., J.R.-H., R.R.-V., C.S., M.S., E.S., F.S.-F., D.S., K.S., V.G.T., N.J.T., M.R.T., S.T.-M., R.P.V., A.W., and J.S.Y. analyzed the genomes. J.M.V.-E., E.M.L., S.L.-G., F.S.-F., R.R.-V., S.T.-M., V.G., L.M.C., and A.H.-E performed the transcriptomic analysis. S.P., S.B., K.M., and A.I. performed the genetic characterization of *madI*. L.M.C., B.A.H., S.T.-M, A.I., A.H.-E., T.G., and I.V.G. wrote the paper with input from all authors.

ACKNOWLEDGMENTS

This article is dedicated to the memory of Max Delbrück (1906–1981). The work by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The project was supported by European funds (European Regional Development Fund, ERDF), the Spanish Ministerio de Economía y Competitividad (BIO2005-25029-E, BIO2015-67148-R), and the Regional Government (Junta de Andalucía, P06-CVI-01650) to L.M.C.; Conacyt (Mexico) (FORDECYT-2012-02-193512) to A.H.-E.; the US National Science Foundation (MCB-0920581) to A.I.; and the Czech Science Foundation (13-33039S) to M.E. We thank A. Gryganskyi, D. Armaleo, and P. Dyer for allowing us to use the genome sequences from *Lichtheimia corymbifera*, *Cladonia grayi*, and *Xanthoria parietina*, respectively, prior to publication. We thank M.M. Gil Sánchez for *P. blakesleeanus* photographs in Figure 1A.

Received: December 23, 2015
Revised: March 22, 2016
Accepted: April 13, 2016
Published: May 26, 2016

REFERENCES

1. Cerdá-Olmedo, E. (2001). *Phycomyces* and the biology of light and color. FEMS Microbiol. Rev. *25*, 503–512.
2. Corrochano, L.M., and Garre, V. (2010). Photobiology in the Zygomycota: multiple photoreceptor genes for complex responses to light. Fungal Genet. Biol. *47*, 893–899.
3. Ma, L.J., Ibrahim, A.S., Skory, C., Grabherr, M.G., Burger, G., Butler, M., Elias, M., Idnurm, A., Lang, B.F., Sone, T., et al. (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. PLoS Genet. *5*, e1000549.
4. Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708–713.
5. Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pöhlmann, R., Luedi, P., Choi, S., et al. (2004). The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science *304*, 304–307.
6. Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature *428*, 617–624.
7. Sanz, C., Rodríguez-Romero, J., Idnurm, A., Christie, J.M., Heitman, J., Corrochano, L.M., and Eslava, A.P. (2009). *Phycomyces* MADB interacts with MADA to form the primary photoreceptor complex for fungal phototropism. Proc. Natl. Acad. Sci. USA *106*, 7095–7100.
8. Idnurm, A., Rodríguez-Romero, J., Corrochano, L.M., Sanz, C., Iturriaga, E.A., Eslava, A.P., and Heitman, J. (2006). The *Phycomyces madA* gene encodes a blue-light photoreceptor for phototropism and other light responses. Proc. Natl. Acad. Sci. USA *103*, 4546–4551.
9. Silva, F., Torres-Martínez, S., and Garre, V. (2006). Distinct *white collar-1* genes control specific light responses in *Mucor circinelloides*. Mol. Microbiol. *61*, 1023–1037.
10. McLaughlin, D.J., Hibbett, D.S., Lutzoni, F., Spatafora, J.W., and Vilgalys, R. (2009). The search for the fungal tree of life. Trends Microbiol. *17*, 488–497.
11. Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E.G., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature *452*, 88–92.
12. Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. USA *108*, 9166–9171.

13. Huerta-Cepas, J., and Gabaldón, T. (2011). Assigning duplication events to relative temporal scales in genome-wide studies. Bioinformatics *27*, 38–45.
14. Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLoS Biol. *13*, e1002220.
15. Marcet-Houben, M., Marceddu, G., and Gabaldón, T. (2009). Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. BMC Evol. Biol. *9*, 295.
16. Schwartze, V.U., Winter, S., Shelest, E., Marcet-Houben, M., Horn, F., Wehner, S., Linde, J., Valiante, V., Sammeth, M., Riege, K., 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*, e1004496.
17. Lafon, A., Han, K.H., Seo, J.A., Yu, J.H., and d'Enfert, C. (2006). G-protein and cAMP-mediated signaling in aspergilli: a genomic perspective. Fungal Genet. Biol. *43*, 490–502.
18. Li, L., Wright, S.J., Krystofova, S., Park, G., and Borkovich, K.A. (2007). Heterotrimeric G protein signaling in filamentous fungi. Annu. Rev. Microbiol. *61*, 423–452.
19. Müller, P., Leibbrandt, A., Teunissen, H., Cubasch, S., Aichinger, C., and Kahmann, R. (2004). The Gbeta-subunit-encoding gene *bpp1* controls cyclic-AMP signaling in *Ustilago maydis*. Eukaryot. Cell *3*, 806–814.
20. Wang, P., Perfect, J.R., and Heitman, J. (2000). The G-protein β subunit GPB1 is required for mating and haploid fruiting in *Cryptococcus neoformans*. Mol. Cell. Biol. *20*, 352–362.
21. Xue, C., Hsueh, Y.P., and Heitman, J. (2008). Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. FEMS Microbiol. Rev. *32*, 1010–1032.
22. Idnurm, A., Verma, S., and Corrochano, L.M. (2010). A glimpse into the basis of vision in the kingdom Mycota. Fungal Genet. Biol. *47*, 881–892.
23. Peñalva, M.A., Tilburn, J., Bignell, E., and Arst, H.N., Jr. (2008). Ambient pH gene regulation in fungi: making connections. Trends Microbiol. *16*, 291–300.
24. Nicolás, F.E., Moxon, S., de Haro, J.P., Calo, S., Grigoriev, I.V., Torres-Martínez, S., Moulton, V., Ruiz-Vázquez, R.M., and Dalmay, T. (2010). Endogenous short RNAs generated by Dicer 2 and RNA-dependent RNA polymerase 1 regulate mRNAs in the basal fungus *Mucor circinelloides*. Nucleic Acids Res. *38*, 5535–5541.
25. Ruiz-Roldán, M.C., Garre, V., Guarro, J., Mariné, M., and Roncero, M.I. (2008). Role of the white collar 1 photoreceptor in carotenogenesis, UV resistance, hydrophobicity, and virulence of *Fusarium oxysporum*. Eukaryot. Cell *7*, 1227–1230.
26. Idnurm, A., and Heitman, J. (2005). Light controls growth and development via a conserved pathway in the fungal kingdom. PLoS Biol. *3*, e95.
27. Lee, S.C., Li, A., Calo, S., and Heitman, J. (2013). Calcineurin plays key roles in the dimorphic transition and virulence of the human pathogenic zygomycete *Mucor circinelloides*. PLoS Pathog. *9*, e1003625.
28. Lipson, E.D., and Terasaka, D.T. (1981). Photogeotropism in *Phycomyces* double mutants. Exp. Mycol. *5*, 101–111.
29. Sanz, C., Benito, E.P., Orejas, M., Alvarez, M.I., and Eslava, A.P. (2010). Protein-DNA interactions in the promoter region of the *Phycomyces carB* and *carRA* genes correlate with the kinetics of their mRNA accumulation in response to light. Fungal Genet. Biol. *47*, 773–781.
30. Larhammar, D., Nordström, K., and Larsson, T.A. (2009). Evolution of vertebrate rod and cone phototransduction genes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 2867–2880.
31. Campuzano, V., Díaz-Mínguez, J.M., Eslava, A.P., and Alvarez, M.I. (1990). A new gene (*madI*) involved in the phototropic response of *Phycomyces*. Mol. Gen. Genet. *223*, 148–151.
32. Alvarez, M.I., Eslava, A.P., and Lipson, E.D. (1989). Phototropism mutants of *Phycomyces blakesleeanus* isolated at low light intensity. Exp. Mycol. *13*, 38–48.
33. Chaudhary, S., Polaino, S., Shakya, V.P., and Idnurm, A. (2013). A new genetic linkage map of the zygomycete fungus *Phycomyces blakesleeanus*. PLoS ONE *8*, e58931.

Current Biology, Volume *26*

Supplemental Information

Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication

Luis M. Corrochano, Alan Kuo, Marina Marcet-Houben, Silvia Polaino, Asaf Salamov, José M. Villalobos-Escobedo, Jane Grimwood, M. Isabel Álvarez, Javier Avalos, Diane Bauer, Ernesto P. Benito, Isabelle Benoit, Gertraud Burger, Lola P. Camino, David Cánovas, Enrique Cerdá-Olmedo, Jan-Fang Cheng, Angel Domínguez, Marek Eliáš, Arturo P. Eslava, Fabian Glaser, Gabriel Gutiérrez, Joseph Heitman, Bernard Henrissat, Enrique A. Iturriaga, B. Franz Lang, José L. Lavín, Soo Chan Lee, Wenjun Li, Erika Lindquist, Sergio López-García, Eva M. Luque, Ana T. Marcos, Joel Martin, Kevin McCluskey, Humberto R. Medina, Alejandro Miralles-Durán, Atsushi Miyazaki, Elisa Muñoz-Torres, José A. Oguiza, Robin A. Ohm, María Olmedo, Margarita Orejas, Lucila Ortiz-Castellanos, Antonio G. Pisabarro, Julio Rodríguez-Romero, José Ruiz-Herrera, Rosa Ruiz-Vázquez, Catalina Sanz, Wendy Schackwitz, Mahdi Shahriari, Ekaterina Shelest, Fátima Silva-Franco, Darren Soanes, Khajamohiddin Syed, Víctor G. Tagua, Nicholas J. Talbot, Michael R. Thon, Hope Tice, Ronald P. de Vries, Ad Wiebenga, Jagjit S. Yadav, Edward L. Braun, Scott E. Baker, Victoriano Garre, Jeremy Schmutz, Benjamin A. Horwitz, Santiago Torres-Martínez, Alexander Idnurm, Alfredo Herrera-Estrella, Toni Gabaldón, and Igor V. Grigoriev

Supplemental Figures and Tables

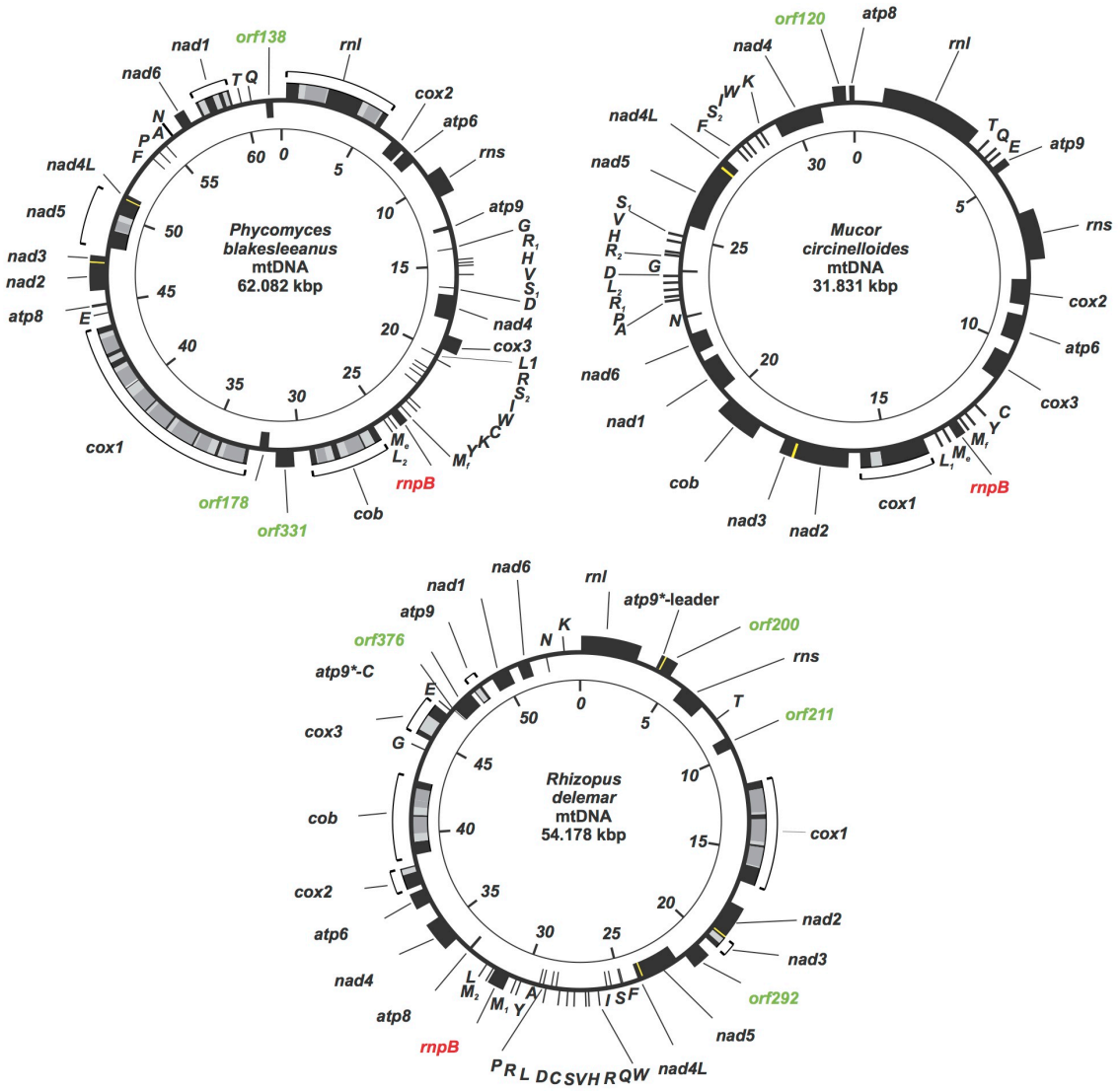

Figure S1 (related to Figure 1). Gene maps of Mucoromycotina mitochondrial DNAs.
Black boxes represent genes. Transfer RNAs are indicated by the one-letter code of their cognate amino acid. Genes with several copies are distinguished by subscript numbers. Me, Mf, genes coding for elongator and initiator tRNAMet, respectively. Gene-name colors depict taxonomic gene distribution. Black, common genes found in most animal, fungal, and protist mtDNAs; grey portions represent introns, with intronic open reading frames (ORFs) shown in dark grey. Red, expanded structural RNA gene set mostly present in mtDNAs of protists and plants and rarely in fungi. Green, hypothetical protein-coding genes that are unique to a given organism; the number specifies the amino acid count in the ORF. Genes on the outer circle are transcribed in a clockwise direction, those on the inner circle are transcribed counter clockwise. The innermost circle serves as size marker.

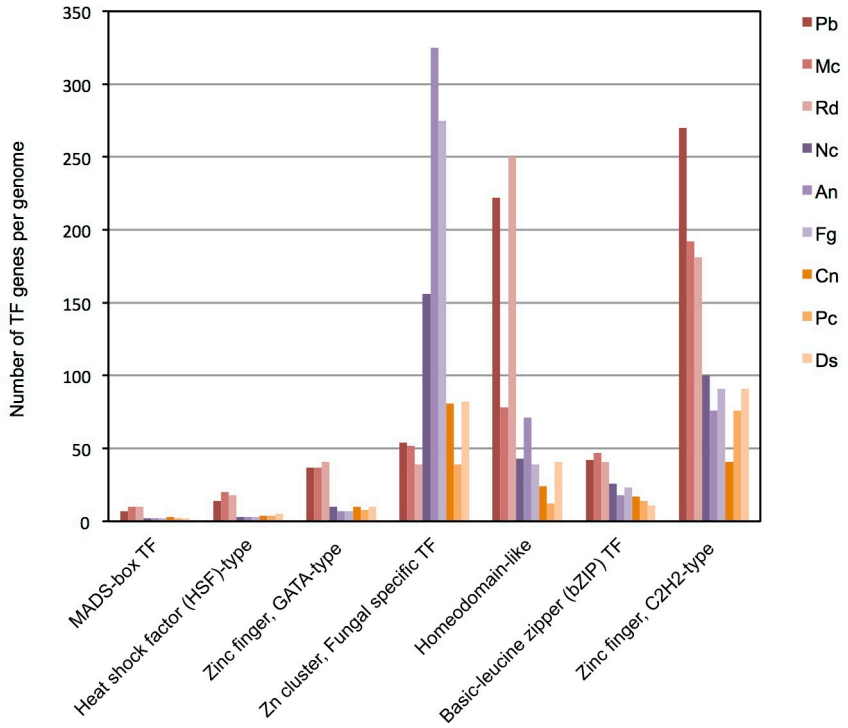

Figure S2 (related to Figure 2). Expanded and under-represented families of transcription factors in *P. blakesleeanus*, *M. circinelloides* and other fungi.
The plot shows the number of transcription factors (TF) for each family in representative fungi of the Mucoromycotina, Ascomycota and Basidiomycota fungi. Pb, *Phycomyces blakesleeanus*; Mc, *Mucor circinelloides*; Rd, *Rhizopus delemar*; Nc, *Neurospora crassa*; An, *Aspergillus nidulans*; Fg, *Fusarium graminearum*; Cn, *Cryptococcus neoformans*; Pc, *Phanerochaete chrysosporium*; Ds, *Dichomitus squalens*.

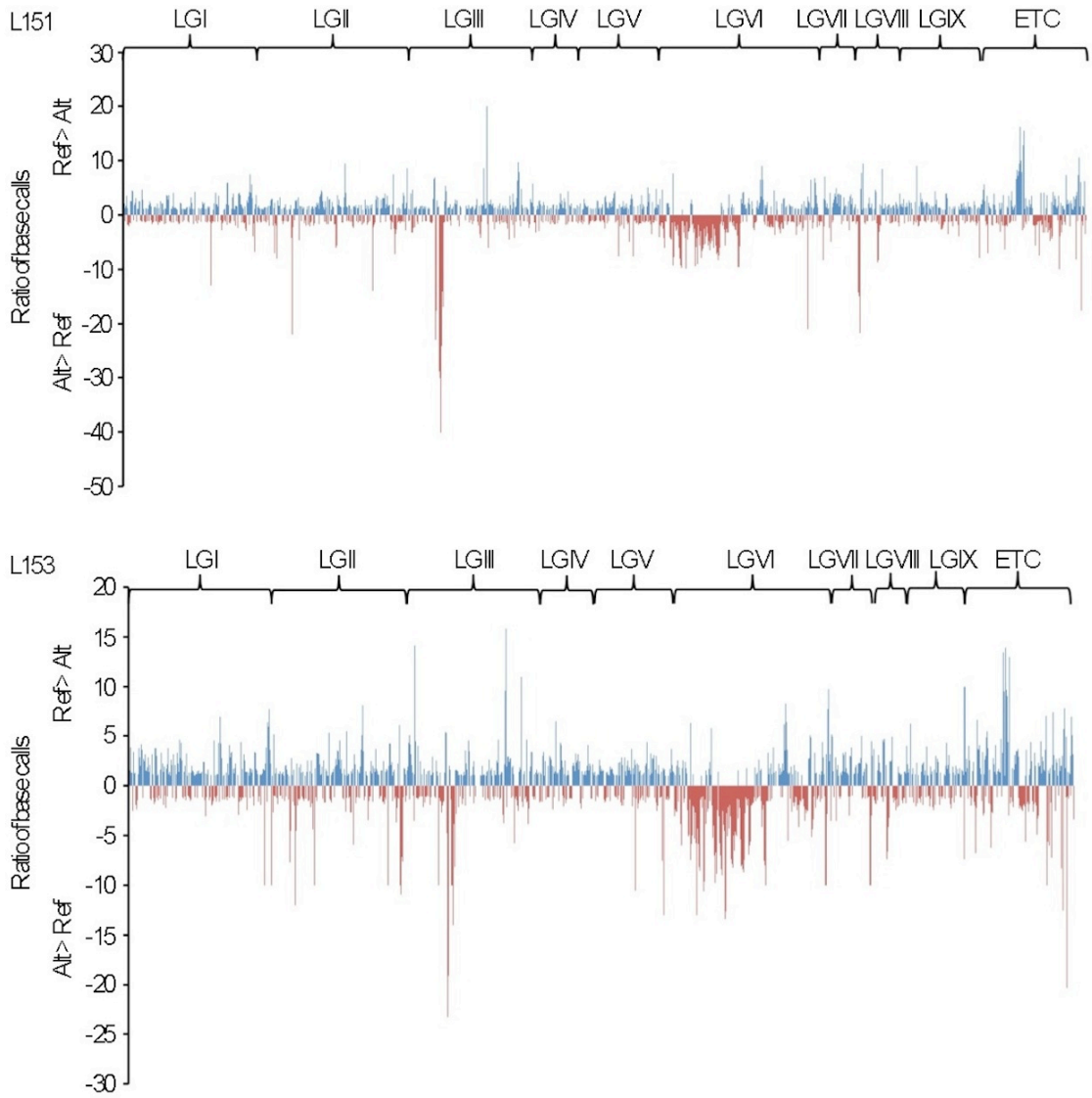

Figure S3 (related to Figure 4). Heterozygosity in the *madI* strains L151 and L153.
The graphs show the distribution of heterozygous SNPs as the ratio of the number of reads specifying the most numerically prevalent base call to the number of reads specifying the alternate base call. The horizontal axis corresponds to the distribution along linkage groups and is not to scale. The vertical axis shows the ratio of the alternate base calls with positive (blue) values indicating that the reference genome base call is more frequent and negative (red) values indicating that the alternate base call is more frequent. LG= Linkage Group, ETC = scaffolds not assigned to linkage groups.

Table S1 (related to Figure 1). Main features of the *P. blakesleeanus* and *M. circinelloides* genomes.

S1A. Main features of the *P. blakesleeanus* and *M. circinelloides* genome assemblies

Nuclear Genome Assembly	***Phycomyces***	***Mucor***
Main genome scaffold total	80	26
Main genome contig total	350	26
Main genome scaffold sequence total	53.9 Mbp	36.6 Mbp
Main genome contig sequence total	53.4 Mbp	36.6 Mbp
Estimated % sequence bases in gaps	1.1%	0.0%
Main genome scaffold N50 / L50	11/1.5 Mbp	4/4.3 Mbp
Main genome contig N50 / L50	41/370.4 kbp	4/4.3 Mbp
Number of scaffolds >50 Kbp	51	15
% main genome in scaffolds >50	99.4%	99.7%
% ESTs aligned to scaffolds	98%	95%

S1B. Genomic libraries included in the *P. blakesleeanus* and *M. circinelloides* whole genome assembly and their respective assembled sequence coverage levels.

	Library Type	**Average insert size, pb**	**Number of reads**	**Sequence coverage (x)**
Phycomyces	Sanger, 3kb	2,834	329,772	2.59
	Sanger, 8kb	6,474	400,608	4.29
	Sanger, fosmid1	35,155	27,264	0.26
	Sanger, fosmid2	35,230	46,080	0.35
	Total		803,724	7.49
Mucor	Sanger, 3kb	2,527	216,830	3.60
	Sanger, 8kb	6,516	340,704	5.02
	Sanger, fosmid	35,117	82,175	0.87
	Total		639,709	9.49

S1C. *P. blakesleeanus* and *M. circinelloides* filtered models classified by gene prediction method.

Prediction method	***Phycomyces* gene models**	***Mucor* gene models**
Total	16,528	11,719
Ab initio	8,494 (51%)	4,839 (41%)
Protein-based	7,261 (44%)	4,901 (41%)
EST-based	773 (5%)	1,979 (17%)

S1D. Properties of *P. blakesleeanus*, *M. circinelloides*, and *R. delemar* gene models.

Property or number	***Phycomyces***	***Mucor***	***Rhizopus***
Avg. gene length	1628 nt	1429 nt	1212 nt
Avg. transcript length	1128 nt	1176 nt	1028 nt
Avg. protein length	353 aa	379 aa	343 aa
Avg. exon length	249 nt	311 nt	310 nt
Avg. intron length	143 nt	93 nt	81 nt
Avg. exon frequency/gene	4.5	3.8	3.3
# multiexon genes	14640 (89%)	9701 (83%)	13540 (78%)
Mean CDS GC content	43.3%	46.3%	40.2%
Mean intron GC content	29.3%	40.1%	30.0%
# genes with similarity to protein in nr	11116 (67%)	9597 (82%)	12303 (70%)
# genes with similarity to gene in same genome	12517 (76%)	8366 (72%)	13776 (79%)
# genes with similarity to gene in other two Mucoromycotina	9632 (58%)	8923 (76%)	11658 (67%)
# genes with EST coverage	5461 (33%)	3897 (33%)	ND
# genes with Pfam domain	7191 (44%)	7524 (64%)	7483 (43%)
# genes with signal peptide	1818 (11%)	1498 (13%)	1546 (9%)
# genes with transmembrane domain	2682 (16%)	2101 (18%)	2133 (12%)
# genes with EC number	2328 (14%)	2596 (22%)	2784 (16%)
# genes with GO term	6408 (39%)	6543 (56%)	8452 (48%)

S1E. Repetitive sequence content of the *P. blakesleeanus* and *M. circinelloides* genomes. Values in parentheses represent the percentage of the genome occupied by the respective repeat family.

Transposable Element Family	***Phycomyces***	***Mucor***
Class I	38 (5.4%)	17 (5.3%)
LTR	9	5
LINE	12	8
Class I Unclassified	17	4
Class II	161 (16.7%)	34 (1.7%)
CMC-EnSpm	31	0
TC1 Mariner	29	3
MuLE	19	0
PIF-Harbinger	9	0
Helitron	4	1
Merlin	1	0
Crypton	1	0
Class II Unclassified	67	30
Other Unclassified Repeats	36 (6.0%)	23 (10.4%)
Total Number of Families	235 (28.1%)	129 (17.4%)

S1F. rRNA genes in the genomes of *P. blakesleeanus*, *M. circinelloides*, *R. delemar*, and *N. crassa*.

	5S	**18S**	**5.8S/ITS**	**28-26S**
P. blakesleeanus	8	6	7	5
M. circinelloides	5	3	3	3
R. delemar	2	2	3	2
N. crassa	79	>40	43	≈ 52

S1G. EST matches to TE families in the *P. blakesleeanus* and *M. circinelloides* genomes.

Repeat Family	***Phycomyces***	***Mucor***
Class I		
LTR	7/9	2/5
LINE	6/12	5/8
Class I Unclassified	11/17	2/4
Class II		
CMC-EnSpm	24/31	0
TC1 Mariner	11/29	0/3
MuLE	6/19	0
PIF-Harbinger	3/9	0
Helitron	2/4	1/1
Merlin	0/1	0
Crypton	0/1	0
Class II Unclassified	17/67	4/30
Other Unclassified Repeats	19/36	8/23
Total	106/235 (45%)	22/129 (17%)

Table S2 (related to Figure 1). Analysis of duplications in fungal genomes.

S2A. Analysis of duplicated genes in fungal genomes. Number of proteins in gene families (540 MCL gene clusters) and fraction of gene families with more members than average for a selection of fungal genomes. Full species names are provided in the section "Clustering of fungal genes in families" of the Supplemental Experimental Procedures.

		Number of proteins per family	**Families with more members than average (%)**
Mucoromycotina	*P. blakesleeanus*	3.0	54.3
	M. circinelloides	3.2	59.8
	R. delemar	3.6	68.0
	L. corymbifera	2.9	50.2
Mortierellomycotina	*M. alpina*	2.4	43.1
Ascomycota	*S. cerevisiae*	2.0	25.4
	S. pombe	1.9	22.4
	N. crassa	1.9	14.6
	A. nidulans	2.0	16.9
	A. niger	2.1	22.4
	S. sclerotiorum	1.8	13.3
	C. globosum	1.8	13.1
	F. graminearum	2.1	18.3
	P. chrysogenum	2.1	18.9
	M. graminicola	1.9	16.3
	C. heterostrophus	2.0	18.9
	C. grayi	1.9	19.6
	X. parietina	1.8	12.2
	T. melanosporum	1.8	15.9
Basidiomycota	*C. neoformans*	1.7	9.3
	U. maydis	1.7	7.4
	T. mesenterica	1.7	6.9
	S. roseus	1.8	11.1
	P. graminis	1.9	20.2
	M. globosa	1.6	8.3
	P. chrysosporium	2.0	17.0
	W. cocos	2.0	15.7
	T. versicolor	2.0	14.8
	D. squalens	1.9	20.2
	H. annosum	1.9	12.8
	L. bicolor	2.2	22.2
	S. commune	2.0	17.8
	S. lacrymans	1.9	12.6
Chytridiomycota	*B. dendrobatidis*	1.9	16.7

S2B. Predicted duplicated regions (paralogons) in fungal genomes. Weighted number of detected duplicated regions, with number of genes in each duplicated region divided by 3 - the minimum number of orthologs in the paralogon, the P-values for real and maximal observed number of weighed duplicated regions in 1000 simulations, the total length of duplicated regions, and the fraction of duplicated regions to the total genome length using gene-based synteny and gene boundaries to extract coordinates. To identify duplicated regions we requested that at least 10% of genes in duplicated regions were homologous (see Supplemental Experimental Procedures). This value resulted in an overestimation of duplicated DNA for some genomes but allowed the best prediction in a set of duplicated regions in the *S. cerevisiae* genome.

		Number of duplicated regions	**P-value**		**Length of duplicated regions (Mb)**	**Fraction of duplicated regions in the genome (%)**
			Real	**Max**		
Mucoromycotina	*P. blakesleeanus*	123.0	<0.001	7	7.939	14.70
	M. circinelloides	59.7	<0.001	9	2.857	7.80
	R. delemar	624.0	<0.001	8	35.593	77.20
	L. corymbifera	169.3	<0.001	12	8.276	24.60
Mortierellomycotina	*M. alpina*	309.0	<0.001	4	16.564	43.20
Ascomycota	*S. cerevisiae*	156.7	<0.001	11	8.469	70.20
	N. crassa	1.0	0.130	5	0.018	0.04
	A. nidulans	11.3	<0.001	6	0.445	1.50
	A. niger	12.3	<0.001	7	0.765	2.20
	S. sclerotiorum	8,3	<0.001	5	0.279	0.70
	C. globosum	9.0	<0.001	4	0.437	1.30
	F. graminearum	8.7	<0.001	4	0.567	1.60
	P. chrysogenum	17.3	0.001	87	1.441	4.60
	M. graminicola	6.0	<0.001	5	0.648	1.60
	C. heterostrophus	133.7	<0.001	6	5.314	14.60
	C. grayi	48.3	<0.001	4	0.909	2.30
	X. parietina	8.7	<0.001	5	0.628	2.00
	T. melanosporum	3.7	<0.001	2	0.372	0.30
Basidiomycota	*C. neoformans*	5.7	0.003	8	0.458	2.40
	U. maydis	2.0	0.002	3	0.206	1.00
	T. mesenterica	3.3	0.008	5	0.264	0.90
	S. roseus	6.3	0.002	7	0.311	1.50
	P. graminis	252.7	<0.001	5	10.862	12.30
	M. globosa	0	1.000	5	0	0
	P. chrysosporium	10.0	<0.001	9	0.537	1.50
	W. cocos	16.3	<0.001	6	1.281	2.50
	T. versicolor	32.0	<0.001	6	1.378	3.10
	D. squalens	16.0	<0.001	5	0.756	1.80
	H. annosum	35.3	<0.001	5	1.425	4.20
	L. bicolor	180.3	<0.001	5	9.446	15.60
	S. commune	46.0	<0.001	5	1.821	4.70
	S. lacrymans	21.3	<0.001	6	1.007	2.40
Chytridiomycota	*B. dendrobatidis*	22.7	<0.001	6	0.552	2.30

S2C. Predicted duplicated genes in fungal genomes. The table shows the number of duplicated genes in each genome, the fraction of duplicated genes relative to the total number of genes in the genome, the average number of duplicated genes in duplicated regions (paralogons), and the fraction of lineage-species genes in paralogons (*i.e.* genes present only in one species).

		Number of duplicated genes	**Fraction of duplicated genes (%)**	**Average duplicated genes in duplicated regions**	**Lineage-specific genes (%)**
Mucoromycotina	*P. blakesleeanus*	738	4.5	4	39.5
	M. circinelloides	358	3.1	4	17.2
	R. delemar	3744	21.4	13	19.2
	L. corymbifera	1016	8.2	5	27.6
Mortierellomycotina	*M. alpina*	1854	12.7	6	51.3
Ascomycota	*S. cerevisiae*	940	16.0	8	36.6
	N. crassa	6	0.1	3	46.1
	A. nidulans	68	0.6	3	5.0
	A. niger	74	0.7	3	16.9
	S. sclerotiorum	50	0.3	4	85.5
	C. globosum	54	0.5	5	46.4
	F. graminearum	52	0.4	3	43.6
	P. chrysogenum	104	0.9	4	40.0
	M. graminicola	36	0.3	4	75.0
	C. heterostrophus	802	6.0	10	45.8
	C. grayi	290	2.5	3	30.7
	X. parietina	52	0.5	3	25.8
	T. melanosporum	22	0.3	4	31.8
Basidiomycota	*C. neoformans*	34	0.5	4	65.6
	U. maydis	12	0.2	3	61.5
	T. mesenterica	20	0.2	3	63.0
	S. roseus	38	0.7	4	55.1
	P. graminis	1516	7.4	7	73.3
	M. globosa	0	0	0	-
	P. chrysosporium	60	0.6	3	40.8
	W. cocos	98	0.8	4	30.3
	T. versicolor	192	1.3	4	34.8
	D. squalens	96	0.8	3	24.9
	H. annosum	212	1.6	5	43.6
	L. bicolor	1082	4.7	7	67.9
	S. commune	276	1.9	5	59.7
	S. lacrymans	128	1.0	4	38.9
Chytridiomycota	*B. dendrobatidis*	136	1.5	4	78.0

S2D. Duplicated genes in phylogenies. Number of duplicated genes in phylogenies between Mucoromycotina fungi (M), and non-Mucoromycotina fungi (nonM) after individually comparing each Mucoromycotina genome with each of 28 genomes from Dikarya. Mucoromycotina genomes are statistically significantly enriched in 2(M):1(nonM) triplets relative to 2(nonM):1(M) triplets (Pearson chi-square<0.001).

	2(M):1(nonM)	**2(nonM):1(M)**
P. blakesleeanus	4181	1139
M. circinelloides	6210	1115
R. delemar	7637	893
L. corymbifera	5928	1076

Supplemental Experimental Procedures

Fungal strains, cultures, and crosses
The strains of *P. blakesleeanus* and *M. circinelloides* are listed below. Cultures of *P. blakesleeanus* were inoculated on minimal medium agar plates and grown at 22 °C in the dark for two days, exposed to light during 30 min or kept in the dark as control [S1]. To isolate RNAs from sporangiophores, cultures were initiated as described above and were grown for two days in the dark at 22 ºC. Then mycelium was exposed to light for two minutes to induce sporangiophore initiation and returned to the dark for another day. Sporangiophores were then either exposed to light for 30 minutes or kept in the dark, removed, and stored at -80 ºC before RNA extraction. Cultures of *M. circinelloides* were inoculated on cellophane sheets on solid YNB medium and grown for 18 hours at 26 °C in the dark, exposed to light for 20 minutes or kept in the dark as control [S2]. Crosses were performed with *P. blakesleeanus* [S3], and the efficiency of phototropism was measured by the bending of sporangiophores to unilateral white light. We selected a light intensity that allowed us to distinguish the phototropic responses of the wild type from the *madA* and *madI* mutants using the detailed characterization of phototropism in these strains performed by Álvarez et al. [S4].

Strains used in this work.
The *P. blakesleeanus mad* mutations were all induced using chemical mutagenesis.

Species and name	Genotype
Phycomyces blakesleeanus	
NRRL1555/ FGSC 10004	Wild type (−)
UBC21/ FGSC 10459	Wild type (+)
A893/ FGSC 25052	*madA403* (−)
A905/ FGSC 10435	*madC406* (−)
A909 / FGSC 25053	*madJ407* (−)
B2/ FGSC 10430	*madC452* (−)
B16	*madD462* (−)
C6 / FGSC 10450	*carRA12 madF48* (−)
C21/ FGSC 10434	*madA7 pde*-1 (−)
C47 / FGSC 25066	*madA35* (−)
C68 / FGSC 10447	*madD59* (−)
C107	*madD99 madG131* (−)
C109 / FGSC 10432	*madB101* (−)
C110 / FGSC 10449	*madE102* (−)
C149 / FGSC 10448	*madD120* (−)
C307 / FGSC 10451	*madG131* (−)
L51 / FGSC 25074	*madA7 madB103* (−)
L151 / FGSC 10453	*madI714* (−)
L153 / FGSC 10454	*madI716* (−)
L157	*madE720* (−)
L161	*madF724* (−)
L163	*madE726* (−)
Mucor circinelloides	
CBS277.49	Wild type

DNA and RNA purification
Genomic DNA from *P. blakesleeanus* or *M. circinelloides* was purified from mycelia [S5, S6]. RNA was isolated from *P. blakesleeanus* or *M. circinelloides* using the Perfect RNA eukaryotic mini kit (Eppendorf) or the RNeasy Plant Mini Kit (Qiagen). Poly A+ RNA was isolated from total RNA using the Absolutely mRNA Purification kit (Stratagene).

Sequencing and assembly
The genomes of *P. blakesleeanus* and *M. circinelloides* were sequenced using Sanger sequencing on ABI 3730XL capillary machines, using three libraries with different insert length that were sequenced from both ends (Table S1). Totals 639,709 reads for *M. circinelloides* (9.49x coverage) and 803,624 reads (7.49x coverage) for *P. blakesleeanus* were collected. Reads were assembled using a modified version of Arachne [S7] v.20071016 with parameters maxcliq1=100, correct1_passes=0 and BINGE_AND_PURGE=True. The resulting whole genome shotgun assemblies were then used as the basis for finishing the genomes. The genomes of the *P. blakesleeanus mad* mutant strains were sequenced using Illumina technology. cDNA synthesis and cloning was a modified procedure based on the "SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning" (Invitrogen). Expressed Sequence Tags (ESTs) were processed through the JGI EST pipeline.

Genome sequence improvement
To perform genome sequence improvement, the *M. circinelloides* and *P. blakesleeanus* whole genome shotgun assemblies were broken down into scaffolds and each scaffold reassembled with phrap. These scaffold were then improved using our Phred/Phrap/Consed pipeline. Initially all low quality regions and gaps were targeted with computationally selected sequencing reactions completed with 4:1 BigDye terminator: dGTP chemistry (Applied Biosystems). These automated rounds included walking on plasmid subclones using custom primers. Following completion of the automated rounds, a trained finisher manually inspected each assembly. This examination included a visual examination of subclone paired ends and visual inspection of high quality discrepancies and all remaining low-quality areas. Further reactions were then manually selected to improve the genome sequence. These reactions included additional custom primer walks on plasmid subclones and fosmids. The reactions were completed using 4:1 BigDye terminator:dGTP chemistry. Smaller repeats in the sequence were resolved by transposon-hopping 8 kb plasmid clones. Fosmid clones were shotgun sequenced and finished to fill large gaps and resolve larger repeats.

cDNA library construction and sequencing
Poly A+ RNA was isolated from total RNA using the Absolutely mRNA Purification kit (Stratagene). cDNA synthesis and cloning was a modified procedure based on the "SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning" (Invitrogen). 1-2 µg of poly A+ RNA, reverse transcriptase SuperScript II (Invitrogen) and oligo dT-NotI primer (5' GACTAGTTCTAGATCGCGAGCGGCCGCCCT15VN 3') were used to synthesize first strand cDNA. Second strand synthesis was performed with *E. coli* DNA ligase, polymerase I, and RNaseH followed by end repair using T4 DNA polymerase. The SalI adaptor (5' TCGACCCACGCGTCCG and 5' CGGACGCGTGGG) was ligated to the cDNA, digested with NotI (New England Biolabs), and subsequently size selected by gel electrophoresis (1.1% agarose). The cDNA inserts were directionally ligated into the SalI and NotI digested vector pCMVsport6 (Invitrogen). The ligation was transformed into ElectroMAX T1 DH10B cells (Invitrogen).

Library quality was first assessed by randomly selecting 24 clones and PCR amplifying the cDNA inserts with the primers M13-F (5' GTAAAACGACGGCCAGT) and M13-R (5' AGGAAACAGCTATGACCAT) to determine the fraction of insertless clones. Colonies from each library were plated onto agar plates (254 mm plates from Teknova) at a density of approximately 1000 colonies per plate. Plates were grown at 37 °C for 18 hours, then individual colonies were picked and each used to inoculate a well containing LB media with appropriate antibiotic in a 384 well plate (Nunc). Clones in 384 well plates were grown at 37 °C for 18 hours. Plasmid DNA for sequencing was produced by rolling circle amplification [S8] (TempliPhi, GE Healthcare). Subclone inserts were sequenced from both ends using primers complimentary to the flanking vector sequence (Fw: 5'ATTTAGGTGACACTATAGAA, Rv: 5' TAATACGACTCACTATAGGG) and Big Dye terminator chemistry then run on ABI 3730 instruments (Applied Biosystems).

EST sequence processing and assembly
Expressed Sequence Tags (ESTs) were processed through the JGI EST pipeline. ESTs were generated in pairs, a 5' and a 3' end read from each cDNA clone. To trim vector and adaptor sequences, common sequence patterns at the ends of ESTs were identified and removed using an internally developed tool. Insertless clones were identified if either of the following criteria were met: >200 bases of vector sequence at the 5' end or less than 100 bases of non-vector sequence remained. ESTs were then trimmed for quality using a sliding window trimmer (window = 11 bases). Once the average quality score in the window was below the threshold (Q15) the EST was split and the longest remaining sequence segment was retained as the trimmed EST. EST sequences with less than 100 bases of high quality sequence were removed. ESTs were evaluated for the presence of polyA or polyT tails (which, if present, were removed) and the EST reevaluated for length, removing ESTs with less than 100 bases remaining. ESTs consisting of more than 50% low complexity sequence were also removed from the final set of "good ESTs". In the case of resequencing the same EST, the longest high quality EST was retained. Sister ESTs (end pair reads) were categorized as follows: if one EST was insertless or a contaminant then by default

the second sister was categorized as the same. However, each sister EST was treated separately for complexity and quality scores. Finally, EST sequences were compared to the GenBank nucleotide database in order to identify contaminants; non-desirable ESTs such as those matching non-cellular and rRNA sequences were removed.

For clustering, ESTs were evaluated with malign, a kmer based alignment tool, which clusters ESTs based on sequence overlap (kmer = 16, seed length requirement = 32 alignment ID>=98%). Clusters of ESTs were further merged based on sister ESTs using double linkage. Double linkage requires that 2 or more matching sister ESTs exist in both clusters to be merged. EST clusters were then each assembled using CAP3 [S9] to form consensus sequences. Clusters may have more than one consensus sequence for various reasons to include; the clone has a long insert, clones are splice variants or consensus sequences are erroneously not assembled. Cluster singlets are clusters of one EST, whereas CAP3 singlets are single ESTs which had joined a cluster but during cluster assembly were isolated into a separate singlet consensus sequence. ESTs from each separate cDNA library were clustered and assembled separately and subsequently the entire set of ESTs for all cDNA libraries were clustered and assembled together. For cluster consensus sequence annotation, the consensus sequences were compared to SwissProt using BLASTx.

Genome annotation and sequence analysis

Genome assemblies of *P. blakesleeanus* NRRL1555 and *M. circinelloides* CBS277.49 were annotated using the JGI Annotation Pipeline and a subset of genes were curated manually. This led to a filtered set of 16,528 *P. blakesleeanus* and 11,719 *M. circinelloides* gene models with their properties and support by different lines of evidence, summarized in Table S1. Repetitive DNA was annotated using the TEdenovo pipeline and manual curation. All predicted gene models were functionally annotated by the JGI Annotation Pipeline against highly curated databases. Automated mtDNA annotation was performed with MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) and validated manually, in particular regarding intron-exon boundaries and termini of ribosomal RNA genes.

The genome assemblies of *P. blakesleeanus* NRRL1555 and *M. circinelloides* CBS277.49 were annotated using the JGI Annotation Pipeline, which combines several gene predictors: A) protein-based gene models were predicted using FGENESH+ [S10] and GeneWise [S11] seeded by BLASTx alignments of genomic sequence against sequences from the NCBI non-redundant protein set nr, B) ab initio gene models were predicted using GeneMark [S12] and FGENESH, the latter trained on the set of putative full-length genes and reliable protein-based models, and C) transcriptome-based gene models were derived by assembling transcript sequences which were then modelled on genomic sequence. GeneWise models were completed using scaffold data to find start and stop codons. Transcriptome alignments to the genome were used to verify, complete, and extend the gene models. Because multiple gene models per locus were often generated, a single representative gene model for each locus was chosen based on protein similarity and transcriptome support, and used for further analysis. This led to a filtered set of 16,528 *P. blakesleeanus* and 11,719 *M. circinelloides* gene models with their properties and support by different lines of evidence summarized in Table S1. A fraction of the gene models were manually curated by scientist experts in each gene category. The manually curated gene models can be accessed at the JGI genome database.

All predicted gene models were functionally annotated by the JGI Annotation Pipeline using InterProScan [S13] and hardware-accelerated double-affine Smith-Waterman alignments (http://www.timelogic.com/) against highly curated databases such as SwissProt [S14], KEGG [S15], and Pfam [S16]. KEGG hits were used to map EC numbers [S17], and InterPro, KEGG, and SwissProt hits were used to map GO terms [S18]. In addition, predicted proteins were annotated according to KOG classification [S19]. Protein targeting predictions were made with signalP [S20] and TMHMM [S21]. Finally, all proteins of each single genome were aligned by BLASTp to proteins in the GenBank nr database and to each other; after the latter procedure the alignment scores were used as a distance metric for clustering by MCL (http://www.micans.org/mcl/) into a first draft of 2,117 *P. blakesleeanus* and 2,066 *M. circinelloides* multigene families. The same method was used to group the *P. blakesleeanus* and *M. circinelloides* proteins with each other as well as with those of *R. delemar* strain 99-880, which were downloaded from the Broad Institute (http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/) on April 21, 2006. For comparative purposes, the Broad Institute's *R. delemar* supercontigs and genes were imported into the JGI database and were functionally annotated by the JGI Annotation Pipeline in the manner described for *P. blakesleeanus* and *M. circinelloides*.

Based on results from BLAST on the 18 genomes used in the phylome analyses, we divided the Mucoromycotina genes into the following categories: ancestral, fungal-specific, Mucoromycotina-specific or species-specific (Figure 1). Mucoromycotina species contained an average of 5014 genes that had homologs in the two non-fungal out-groups, deeming them of ancestral origin. This number is double the average found in the other fungal species (2320 on average), likely due to the whole genome duplications that occurred in this group. A similar trend is observed in genes of fungal origin (2737 genes versus 1381). Very few of those genes are

found in all of the fungal species: this is mainly due to the presence of microsporidian genomes in our dataset and partly because most of the widespread genes were also found in the outgroups. An average of 840 genes were shared between all four Mucoromycotina species but this number increased to 2568 if we consider all genes that were at least in two members of this group. Surprisingly, despite the fact that *R. delemar* underwent a species-specific whole genome duplication, *P. blakesleeanus* is the Mucoromycotina species with the largest number of species-specific genes (5183). On the contrary, *M. circinelloides* only contained 1867 genes that could only be found in this species.

Genome resequencing

For each strain, genomic DNA (1µg) was sheared into ~250bp fragments using the Covaris E210 (Covaris). The DNA fragments were subjected to end repair, A-tailing, and ligation of Illumina compatible adapters (IDT). The final library was enriched with 10 cycles of PCR (NEB). The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina GA sequencing platform utilizing a paired-end cluster generation kit, v4, and Illumina's cBot instrument to generate clustered flowcells for sequencing. Sequencing of the flowcells was performed on the Illumina GAIIx sequencer using SBS sequencing kits, v4, following a 2x76 run recipe. The libraries were sequenced on Illumina GA II sequencers generating 75bp paired end reads. These reads were aligned to the reference genome and putative SNPs and small indels were called using maq-0.7.1 [S22]. Putative structural variants were called using BreakDancer [S23], filtering for a confidence score of >90.

Identification and analysis of repetitive elements

Repetitive DNA was annotated with REPET [S24, S25] using the TEdenovo pipeline included in REPET to detect and classify repeat families. The repeat family consensus sequences were manually curated by performing BLAST searches to the RepBase protein database [S26] and manually validating the repeat classifications. The REPET pipeline TEannot was used to annotate the repeat families. TE consensus sequences were aligned to the EST consensus with megaBLAST, using an e-value threshold of 10^{-5}. rRNA genes were identified with the program RNAmmer [S27].

***P. blakesleeanus* RNAseq and analysis**

Libraries for RNAseq were prepared with RNA from mycelia and sporangiophores of the wild-type and the *madA madB* mutant strains using the SOLiD whole transcriptome kit protocol (Applied Biosystems) [S28]. Three biological replicates for each condition and two technical replicates of the experiment were performed. Sequencing was performed on a SOLiD 4 ABI sequencer. 50-base-pair-long reads were obtained and mapped to the *P. blakesleeanus* genome V2 using bowtie [S29], searching for end-to-end hits with at most three mismatches. Alignment results were recorded in BAM format for further downstream analysis. Read counts per gene were calculated for each library using a shell script in the statistical software R, and collapsed into a table considering both technical replicas.

Only genes with at least three reads per million were considered for differential expression analyses; this was done using the edgeR package [S30]. Normalization of the data was performed using the Cox-Reid adjusted likelihood method. We determined that the WTMD3, WTSD3, WTSL2 and ΔL51SD2 libraries had greater dispersion compared to their respective biological replicas, so it was decided not to consider these data for the final analysis. For determining differential expression between the different comparisons we used the generalized linear model likelihood ratio test. False discovery rates (FDR) were calculated using the Benjamini & Hochberg procedure [S31]. Genes with a FDR<0.05 were considered differentially expressed.

The protein sequences were obtained from the gene catalog of *P. blakesleeanus* V2 genome in FASTA format. BLAST comparison with this file was performed against the non-redundant database (nr) of NCBI using Blast2GO version 2.6.0 (http://www.blast2go.org/), with an E-value $\leq 1 \times 10^{-3}$. GO term assignation was performed using an E-value $\leq 1 \times 10^{-3}$, an annotation score ≤ 40, a GO weight of 5 [S32]. The annotation was improved using the protein domains databases InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) and ANNEX [S33]. All GOs ancestors for each gene were obtained using an R script.

Using the GOseq package of Bioconductor, an enrichment analysis was performed with the hypergeometric distribution method. GO terms with FDR ≤ 0.05 were considered significantly enriched in each comparison. The RNAseq results have been submitted to the GEO database with accession GSE64369.

Microarray analysis of *M. circinelloides* expression

Total RNAs (0.2 µg) from each sample were amplified, labelled and hybridized to a custom Agilent Microarray (Agilent, ID G2509F). RNA labelling was performed according to the Low Input Quick Amp Labeling kit (Agilent). Images of the microarrays were acquired using a G2505C Scanner (Agilent). Raw data were obtained from images using Feature Extraction Software v. 10.7.3.1 (Agilent). All statistical and differential expression

analyses were carried out with the Limma package from Bioconductor (http://www.bioconductor.org/). Gene expression was considered statistically significant when FDR<0.05.

Clustering of fungal genes in families

The Markov clustering method (MCL) algorithm [S34] was used to cluster proteins from the following 34 fungi with complete genomes. Four Mucoromycotina: *Phycomyces blakesleeanus*, *Mucor circinelloides*, *Rhizopus delemar*, *Lichtheimia corymbifera*; 1 Mortierellomycotina: *Mortierella alpina*. 14 Ascomycota: *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Neurospora crassa*, *Aspergillus nidulans*, *Aspergillus niger*, *Sclerotinia sclerotiorum*, *Chaetomium globosum*, *Fusarium graminearum*, *Penicillium chrysogenum*, *Mycosphaerella graminicola*, *Cochliobolus heterostrophus*, *Cladonia grayi*, *Xanthoria parietina*, *Tuber melanosporum*. 14 Basidiomycota: *Cryptococcus neoformans*, *Ustilago maydis*, *Tremella mesenterica*, *Sporobolomyces roseus*, *Puccinia graminis*, *Malassezia globosa*, *Phanerochaete chrysosporium*, *Wolfiporia cocos*, *Trametes versicolor*, *Dichomitus squalens*, *Heterobasidion annosum*, *Laccaria bicolor*, *Schizophyllum commune*, *Serpula lacrymans*. One Chytridiomycota: *Batrachochytrium dendrobatidis*.

Detection of duplicated regions (paralogons)

Several methods have been developed for the identification of duplicated regions (paralogons) [S35] but, until recently, there was no way to confirm their accuracy. A manually curated set of duplicated regions in the *S. cerevisiae* genome is considered as the standard set for ancient duplicated regions [S36]. We implemented a simple approach for detecting such regions with two parameters: Nmin (minimal number of duplicated gene pairs) and Fhom (minimal fraction of homologous genes in the region). Because possible duplications in the Mucoromycotina may be even more ancient we removed the constraint of ordered genes in the algorithm. When tested on *S. cerevisiae* best accuracy was achieved at Nmin=3 and Fhom=10%, with 58 of the 67 paralogons predicted correctly, which amounts to sensitivity Sn=91% and specificity Sp=87%. At the level of detecting duplicated pairs of genes, the method predicts in total 515 pairs in paralogons, 436 of them correctly (Sn=91%, Sp=85%). We applied the above method to detect paralogons in the fungal genomes described in the previous section. We also computed potential paralogons in 1000 simulated genomes with randomly shuffled order of genes in scaffolds/chromosomes.

Detection of whole genome duplication (WGD) signatures

For each pair of genomes we concentrated only on families consisting of three members where genes with two copies duplicated after the split of the two compared species resulting in "2:1 after split" triplets. Our rationale is that the number of such triplets should be significantly larger in one genome relative to another after the genome had undergone a WGD event. For every pair of compared genomes we first selected the triplets of homologous genes (with BLASTp score at least 100 and alignment covering at least 80% of the compared genes length) and which have no other homologs in both genomes. Then for each selected triplet we constructed a phylogenetic tree and selected triplets for which duplication occurred after the split between the two genomes, *i.e.* duplicated genes are closer to each other than to the homolog in another genome.

Phylome reconstruction

The phylome - *i.e.* the complete collection of phylogenetic trees for each gene in a genome - was reconstructed for the genomes of the four Mucoromycotina species: *R. oryzae*, *P. blakesleeanus*, *M. circinelloides* and *L. corymbifera*. A fifth phylome was reconstructed using the related species, *Mortierella alpina*, as seed. All phylomes were reconstructed in the context of 13 other fungal species: two additional early diverging fungi (*Batrachochytrium dendrobatidis*, *Homoloaphlyctis polyrhiza*), three Microsporidia (*Nematocida parisii*, *Encephalitozoon cuniculi*, *Nosema ceranae*) and six Dikarya species (*Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Puccinia graminis*, *Ustilago maydis*, and *Cryptococcus neoformans*). *Nematostella vectensis* (a sea anemone) and *Monosiga brevicollis* (a choanoflagellate) were used as outgroups.

The phylome was reconstructed using an automated pipeline described previously [S37]. Briefly, for each protein encoded in the three seed genomes a Smith-Waterman search was performed against the proteome database. Results were then filtered using an e-value cut-off of $E<10^{-5}$ and requiring a continuous overlapping region of 0.5 over the query sequence. A limit of 150 BLAST hits for each protein was used. Homologous sequences were then aligned using three different programs: MUSCLE v3.8 [S38], MAFFT [S39] and kalign [S40]. Alignments were performed in forward and reverse direction and the six resulting alignments were combined with M-COFFEE [S41]. TrimAl v1.3 [S42] was then used to trim the alignment using a consistency-score cut-off of 0.1667 and a gap-score cut-off of 0.9. The alignments were then used to reconstruct maximum likelihood trees. The model best fitting the data was selected as follows: for each evolutionary model (JTT, LG, WAG, Blosum62, MtREV, VT and Dayhoff) a Neighbour Joining (NJ) tree was reconstructed as implemented in BioNJ [S43] then the likelihood of this topology was computed, allowing branch-length optimization as implemented in PhyML v3.0 [S44]; the model best fitting the data was determined by the AIC criterion [S45].

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. The trees and alignments reconstructed for the three phylomes can be found at phylomeDB [S46] (http://phylomedb.org), with the phylomeIDs 252, 253, 254, 255 and 256.

Species tree reconstruction
Two methodologies were used to reconstruct the species trees encompassing the 18 species used in the phylome reconstruction. First, a gene concatenation analysis was done based on 49 proteins encoded in genes that were found in single copy in at least 15 out of the 18 species. The concatenated alignment contained 25,023 amino acids. A ML tree was reconstructed using PhyML [S44] with LG model [S47] selected, four rate categories were used and invariant positions were inferred from the data. The second species tree was reconstructed using a super-tree approach. Duptree [S48] was used on the 13,734 trees reconstructed in the *R. delemar* phylome. This super-tree approach aims to find the species tree that minimizes the number of duplications inferred when reconciling genes trees with the species tree. The topologies obtained by both methods were identical and only the ML phylogeny is shown in Fig. 1C.

Mapping duplication rates
Single gene trees were scanned using ETE v2 [S49]. For each tree, duplications were found using a species-overlap algorithm. The duplications were mapped onto the species trees assuming that they occurred at the common ancestor of all the species involved in the duplication node. The number of duplications was then divided by the number of trees that could potentially contain a duplication at that given point. Trees that contained species-specific expansions of more than five members for the seed species were omitted.

Identification of G-protein coupled receptors
A specialized hidden Markov model specific to identify transmembrane (TM) regions of GPCRs, GPCRHMM [S50], was applied. This technique predicts if a given sequence is or is not a GPCR based on different TM topology features, such as loop length and amino acid composition. In a sensitivity-selectivity test, the sensitivity of GPCRHMM was reported to be about 15% higher than that of any other predictor tested, at comparable false positive rates.

To reduce the chances of wrongly classifying proteins as GPCRs, the PHOBIUS method, another TM predictor [S51], was applied to estimate independently the number of TM regions for each putative GPCR. Only in cases where PHOBIUS also predicted seven TM regions was the entry retained. This procedure resulted in the removal of about half the sequences from the initial set of candidates.

The resulting list was classified with three different tools. The first classification used BLASTclust with the following options: 50% coverage, identity cutoff 30% (http://toolkit.tuebingen.mpg.de/). In parallel, the list was searched with a GPCR classification tool [S52], and an alignment and a phylogenetic tree were also constructed directly from the list of candidate sequences. The GPCR-2L procedure may have been too stringent for fungal sequences, excluding, for example, Class X defined in *Trichoderma* spp. Furthermore, a class of GPCR-like proteins with similarity to the PQ-loop-domain containing protein Stm1 of *Schizosaccharomyces pombe* was not detected by GPCRHMM. Stm1 is a predicted 7TM protein by both PHOBIUS and GPCRHMM, however, it fails to meet the criteria for GPCR prediction by GPCRHMM. There is functional evidence that Stm1 is indeed a GPCR, based on a screening assay in yeast [S53, 54]. The homologs of Stm1 were identified by BLASTp in the Mucoromycotina databases. The list in Data S1 combines the results of these analyses, annotated to include the smaller subset identified by GPCR-2L.

Identification of protein kinases
In order to conduct as unbiased a search as possible, pre-calculated HMM profiles (http://www.compbio.dundee.ac.uk/kinomer) [S55] were used to identify and classify all kinases in the predicted proteomes belonging to one of 10 families. A protein was considered to belong to a family for which its raw bit score was the best (often there is multiple family classification) and it is at least ≥ 20.

Identification of cell wall biosynthesis proteins
The study was performed manually using a number of programs. PSORTII (http://www.genscript.com/psort/psort2.html) was used for the prediction of proteins with extracellular location and the presence of a signal peptide. Protein location was further confirmed with an extension of PSORTII (http://www.genscript.com/psort/wolf_psort.html). The presence of a signal peptide was confirmed by SignalP (http://www.cbs.dtu.dk/services/SignalP/). Proteins containing a putative extracellular location and a signal peptide were further analyzed for the presence of serine/threonine residues, and putative glycosylation sites using PROSITE (http://us.expasy.org/tools/scanprosite/). The presence of a GPI motif was analyzed with the big-PI Fungal Predictor algorithm (http://mendel.imp.univie.ac.at/gpi/fungi_server.html). The presence of amino

acid repeats in the proteins was analyzed with the program RADAR (http://www.ebi.ac.uk/Radar/). BLAST analyses were run at the SIB Blast Network service (http://us.expasy.org/tools/blast/) and at http://www.ebi.ac.uk/blastall/. Identification of important motifs in the analyzed proteins was carried out using the program SAPS (http://www.ebi.ac.uk/Tools/seqstats/saps/).

Identification of transcription factors

The genome-wide assignments of TFs were made by scanning the InterProScan [S56] predictions for all known DNA-binding domains (DBDs) assigned to transcriptional regulation. The library of the corresponding DBDs was collected from different sources including Pfam [S57] and DBD [S58] databases and literature. The InterProScan predictions were downloaded from the JGI database (http://jgi.doe.gov). The *P. blakesleeanus* and *M. circinelloides* genomes contain 879 and 650 genes for transcription factors (TFs), respectively (about 5% of the protein-coding genes), with an abundance of C_2H_2 Zn finger TFs. The 879 TFs in the *P. blakesleeanus* genome are assigned to 49 families of DNA-binding domains as based on InterProScan predictions.

Identification of TRAFAC class GTPases and regulators

The sequences were identified by BLASTp and PSI-BLAST searches of the database of predicted *P. blakesleeanus* proteins, *R. delemar* proteins and the nr database at NCBI (the remaining fungi). For *P. blakesleeanus* and *R. delemar*, the genome was further checked by tBLASTn to identify possible homologs that had been annotated incorrectly or not at all. For the remaining fungal genomes analysed, tBLASTn searches were conducted only in cases of suspicious absence of a homolog in the protein database (leading in several cases to identification of an un-annotated gene), hence it is possible that some additional un-annotated genes have been missed. Assignment of mutual orthologs is based mainly on reciprocal BLAST comparisons and in a few cases should be verified by phylogenetic analysis (accession numbers of individual GTPases from Dikarya are available upon request). Unless stated otherwise, the names assigned to groups of orthologs follow the nomenclature of the corresponding *S. cerevisiae* genes. Probable or known function of individual GTPases is assigned by consulting the literature dealing with characterisation of orthologs in fungal or other species. Fungal sequences falling into various categories of known regulators of the Ras GTPase superfamily were identified and analysed by an approach similar to that used for the analysis of fungal GTPases. In addition, for identification of proteins containing poorly conserved regulatory domains, we also used HMMER searches with profile HMMs built from multiple alignments retrieved from Pfam (http://www.sanger.ac.uk/Software/Pfam/) or SMART (http://smart.embl-heidelberg.de/) collections. The number of TRAFAC class GTPases identified in the *P. blakesleeanus* genome (148) is lower than the number for the *R. delemar* genome (192) but higher than the number for any Dikarya genome analysed, 70-133. The total number of Ras superfamily proteins encoded by the *P. blakesleeanus* and *R. delemar* genomes (213 and 255, respectively) is higher than in the Dikarya (from 77 to 100). The higher number of GTPases and Ras superfamily proteins in the Mucoromycotina fungi compared to Dikarya is due to both retention of a higher number of ancestral genes and more extensive lineage-specific gene duplications.

Supplemental references

S1. Idnurm, A., Rodríguez-Romero, J., Corrochano, L.M., Sanz, C., Iturriaga, E.A., Eslava, A.P., and Heitman, J. (2006). The *Phycomyces madA* gene encodes a blue-light photoreceptor for phototropism and other light responses. Proc Natl Acad Sci USA *103*, 4546-4551.

S2. Silva, F., Torres-Martínez, S., and Garre, V. (2006). Distinct *white collar-1* genes control specific light responses in *Mucor circinelloides*. Mol Microbiol *61*, 1023-1037.

S3. Chaudhary, S., Polaino, S., Shakya, V.P., and Idnurm, A. (2013). A new genetic linkage map of the zygomycete fungus *Phycomyces blakesleeanus*. PLoS One *8*, e58931.

S4. Alvarez, M.I., Eslava, A.P., and Lipson, E.D. (1989). Phototropism mutants of *Phycomyces blakesleeanus* isolated at low light intensity. Exp Mycol *13*, 38-48.

S5. Weinkove, D., Poyatos, J.A., Greiner, H., Oltra, E., Avalos, J., Fukshansky, L., Barrero, A.F., and Cerdá-Olmedo, E. (1998). Mutants of *Phycomyces* with decreased gallic acid content. Fungal Genet Biol *25*, 196-203.

S6. Sambrook, J., and Russell, D.W. (2001). Molecular Cloning. A Laboratory Manual, (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

S7. Jaffe, D.B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J.P., Zody, M.C., and Lander, E.S. (2003). Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res *13*, 91-96.

S8. Detter, J.C., Jett, J.M., Lucas, S.M., Dalin, E., Arellano, A.R., Wang, M., Nelson, J.R., Chapman, J., Lou, Y., Rokhsar, D., et al. (2002). Isothermal strand-displacement amplification applications for high-throughput genomics. Genomics *80*, 691-698.
S9. Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Res *9*, 868-877.
S10. Salamov, A.A., and Solovyev, V.V. (2000). Ab initio gene finding in *Drosophila* genomic DNA. Genome Res *10*, 516-522.
S11. Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*, 988-995.
S12. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res *18*, 1979-1990.
S13. Zdobnov, E.M., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847-848.
S14. Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2005). The Universal Protein Resource (UniProt). Nucleic Acids Res *33*, D154-159.
S15. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res *27*, 29-34.
S16. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. (2004). The Pfam protein families database. Nucleic Acids Res *32*, D138-141.
S17. Bairoch, A. (2000). The ENZYME database in 2000. Nucleic Acids Res *28*, 304-305.
S18. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.
S19. Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol *5*, R7.
S20. Nielsen, H., Brunak, S., and von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng *12*, 3-9.
S21. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.
S22. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res *18*, 1851-1858.
S23. Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods *6*, 677-681.
S24. Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol *1*, 166-175.
S25. Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. PLoS One *6*, e16526.
S26. Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet *16*, 418-420.
S27. Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res *35*, 3100-3108.
S28. Vega-Arreguín, J.C., Ibarra-Laclette, E., Jiménez-Moraila, B., Martínez, O., Vielle-Calzada, J.P., Herrera-Estrella, L., and Herrera-Estrella, A. (2009). Deep sampling of the *Palomero* maize transcriptome by a high throughput strategy of pyrosequencing. BMC Genomics *10*, 299.
S29. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
S30. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
S31. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B *57*, 289-300.
S32. Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res *36*, 3420-3435.

S33. Myhre, S., Tveit, H., Mollestad, T., and Laegreid, A. (2006). Additional gene ontology structure for improved biological reasoning. Bioinformatics *22*, 2020-2027.
S34. Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res *30*, 1575-1584.
S35. McLysaght, A., Hokamp, K., and Wolfe, K.H. (2002). Extensive genomic duplication during early chordate evolution. Nat Genet *31*, 200-204.
S36. Gordon, J.L., Byrne, K.P., and Wolfe, K.H. (2009). Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genet *5*, e1000485.
S37. Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L.P., Denisov, I., Kormes, D., Marcet-Houben, M., and Gabaldón, T. (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*, D556-560.
S38. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.
S39. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res *33*, 511-518.
S40. Lassmann, T., and Sonnhammer, E.L. (2005). Kalign--an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics *6*, 298.
S41. Wallace, I.M., O'Sullivan, O., Higgins, D.G., and Notredame, C. (2006). M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res *34*, 1692-1699.
S42. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972-1973.
S43. 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.
S44. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol *59*, 307-321.
S45. Akaike, H. (1975). Information theory and an extension of the maximum likelihood principle. In 2nd International Symposium on Information Theory, B.N. Petrov and F. Csáki, eds. (Budapest, Hungary: Akadémia Kiado), pp. 267–281.
S46. Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L.P., Marcet-Houben, M., and Gabaldón, T. (2014). PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. Nucleic Acids Res *42*, D897-902.
S47. Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. Mol Biol Evol *25*, 1307-1320.
S48. Wehe, A., Bansal, M.S., Burleigh, J.G., and Eulenstein, O. (2008). DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. Bioinformatics *24*, 1540-1541.
S49. Huerta-Cepas, J., Dopazo, J., and Gabaldón, T. (2010). ETE: a python environment for tree exploration. BMC Bioinformatics *11*, 24.
S50. Wistrand, M., Kall, L., and Sonnhammer, E.L. (2006). A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Sci *15*, 509-521.
S51. Kall, L., Krogh, A., and Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. J Mol Biol *338*, 1027-1036.
S52. Xiao, X., Wang, P., and Chou, K.C. (2011). GPCR-2L: predicting G protein-coupled receptors and their types by hybridizing two different modes of pseudo amino acid compositions. Mol Biosyst *7*, 911-919.
S53. Chung, K.S., Won, M., Lee, J.J., Ahn, J., Hoe, K.L., Kim, D.U., Song, K.B., and Yoo, H.S. (2007). Yeast-based screening to identify modulators of G-protein signaling using uncontrolled cell division cycle by overexpression of Stm1. J Biotechnol *129*, 547-554.
S54. Chung, K.S., Won, M., Lee, S.B., Jang, Y.J., Hoe, K.L., Kim, D.U., Lee, J.W., Kim, K.W., and Yoo, H.S. (2001). Isolation of a novel gene from *Schizosaccharomyces pombe*: *stm1+* encoding a seven-transmembrane loop protein that may couple with the heterotrimeric Galpha 2 protein, Gpa2. J Biol Chem *276*, 40190-40201.
S55. Miranda-Saavedra, D., and Barton, G.J. (2007). Classification and functional annotation of eukaryotic protein kinases. Proteins *68*, 893-914.
S56. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236-1240.

S57. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res *42*, D222-230.

S58. Wilson, D., Charoensawan, V., Kummerfeld, S.K., and Teichmann, S.A. (2008). DBD--taxonomically broad transcription factor predictions: new content and functionality. Nucleic Acids Res *36*, D88-92.