

Supporting Information

Methods S1

Sequencing, assembly and annotation

For production of the *M. osmundae* genome, one un-amplified whole-genome shotgun (WGS) and one 4-kb Long Mate Pair (LMP) library were generated as follows. For the WGS library, 1 µg of genomic DNA was sheared to the average of 270 bp in length. The fragments were size selected by SPRI beads followed by end-repair, phosphorylation and A-tailing. Fragments were ligated with HiSeq-compatible adapters (Integrated DNA Technologies, Inc., Coralville, IA, USA) and purified using SPRI beads.

LMP libraries were generated by shearing 15 µg of DNA to approximately 4-kb fragments. The sheared DNA was size selected, end-repaired, and ligated with mate pair adapters containing a LoxP site and Illumina primer sequences. DNA was then circularized with Cre Recombinase and paired-end tags were generated from the circularized DNA by NlaIII restriction digestion. The digested fragments were circularized and amplified by inverse PCR. The prepared sample libraries were quantified using KAPA Biosystems (Woburn, MA, USA) next-generation sequencing library qPCR kit and run on a LightCycler 480 real-time PCR instrument (Roche Applied Science, Indianapolis, IN, USA).

Each of the quantified sample libraries was then prepared for sequencing on the Illumina HiSeq 2000 sequencing platform (Illumina, Inc., San Diego, CA, USA) utilizing a paired-end cluster generation kit, TruSeq PE Cluster Kit v3, and Illumina's cBot instrument to generate clustered flow cells for sequencing. Sequencing of the flow cells was performed on the Illumina HiSeq 2000 sequencer using TruSeq SBS Kit v3, following 2×100 and 1×150 run recipes.

For transcriptome sequencing, poly A RNA was isolated from 10 µg of total RNA using an Absolutely mRNA Purification kit (Agilent Technologies, La Jolla, CA, USA). The isolation procedure was repeated twice to ensure that the sample was free from rRNA contamination. Purified RNA was fragmented using RNA fragmentation reagents (Ambion, Inc.) and used as a template for first-strand cDNA synthesis. Reverse transcription was performed using SuperScript II Reverse Transcription (Invitrogen, Carlsbad, CA, USA), with random hexamer as primer. This was followed by second-strand cDNA synthesis using dNTP mix where dTTP is replaced by dUTP. Double-stranded cDNA fragments were then blunt ended, poly-A tailed and ligated with

Truseq adaptors using the Illumina DNA Sample Prep Kit. Digestion of dUTP was performed using AmpErase UNG (Applied Biosystems, Grand Island, NY, USA); digested cDNA was amplified for 10 cycles by PCR with Illumina Truseq primers and sequenced on the Illumina HiSeq 2000 platform, generating 100-bp paired-end reads.

De novo transcript contig assembly was performed using Rnnotator (Martin *et al.*, 2010) and genome sequence data for both WGS and LMP libraries were assembled using AllPathsLG (release 38445; Gnerre *et al.*, 2011). 22,344 out of 23,482 EST contigs (95.15%) were mapped to the genome assembly with at least 90% identity and 85% coverage. 338 million paired-end 100-bp Illumina reads were used as input and screened for low-quality, low-complexity, adapter-containing and duplicate reads. The remaining 189 million reads were assembled using Velvet v. 1.1.04 (Zerbino & Birney, 2008) with different hash lengths from 53 to 95. Duplicate contigs were removed and contigs with a minimum overlap of 40 bp were merged. To generate a consensus nucleotide sequence, the resulting contigs were extended and polished using the strand information of the reads and single-base error correction by aligning the reads back to each contig with Burrows-Wheeler Aligner (Li & Durbin, 2009).

Sequence data for both WGS and LMP libraries were assembled using AllPathsLG (release 38445; Gnerre *et al.*, 2011). The assembly was run with ploidy = 1 using $80 \times$ input of 4 kb LMP, $90 \times$ coverage of WGS data based on an estimated genome size of 13,500,000 bp and a minimum contig size of 100 bp. Reads were filtered for artifacts, sequence data from mitochondria, one or more N bases and an average phred quality of 30 or higher using kmer analysis. Bases with a q score lower than 20 were trimmed from both ends and reads were screened for a minimum read length (70 bp for the LMP data and 90 bp for the WGS data). NCBI screening did not identify any contamination. 22,344 out of 23,482 EST contigs (95.15%) were mapped to the genome assembly with at least 90% identity and 85% coverage.

Before gene prediction, assembly scaffolds were masked using RepeatMasker (Smit *et al.*, 2010), RepBase library (Jurka *et al.*, 2005), and the most frequent (>150 times) repeats were recognized by RepeatScout (Price *et al.*, 2005). The following combination of gene predictors was run on the masked assembly: *ab initio* Fgenesh (Salamov & Solovyev, 2000) and GeneMark (Ter-Hovhannisyan *et al.*, 2008); homology-based Fgenesh+ (Salamov & Solovyev, 2000) and Genewise (Birney *et al.*, 2004) seeded by BLASTx alignments against the NCBI NR database; and transcriptome-based CombEST (Zhou *et al.*, personal communications). In addition to

protein-coding genes, tRNAs were predicted using tRNAscan-SE (Lowe & Eddy, 1997). All predicted proteins were functionally annotated using SignalP (Nielsen *et al.*, 1997) for signal sequences, TMHMM (Melen *et al.*, 2003) for transmembrane domains, InterProScan (Quevillon *et al.*, 2005) for integrated collection of functional and structured protein domains, and protein alignments to NCBI NR, SwissProt (<http://www.expasy.org/sprot/>), and KEGG (Kanehisa *et al.*, 2006) for metabolic pathways, and KOG (Koonin *et al.*, 2004) for eukaryotic clusters of orthologs. Interpro and SwissProt hits were used to map gene ontology (GO) terms (Ashburner *et al.*, 2000). For each genomic locus, the best representative gene model was selected based on a combination of protein similarity and EST support.

To assess whether the assembly represents a haploid or diploid genome, we attempted to identify syntenic blocks of genes that are present twice in the genome. First, we identified potential diploid pairs of genes whose transcripts form the best bidirectional (reciprocal) Blastn hits with each other and have a pairwise identity > 90%. Thereafter, we selected pairs of genes that form a syntenic block of 2 or more genes on different scaffolds (or 1 gene pair if one or both scaffolds are small and contain only 1 gene). Only two identical regions were detected: S8: 623,143–756,102 and S17: 3–132,962 that were further shown to be an artifact of assembly.

Multi-gene families

Multi-gene families were identified using the JGI clustering pipeline. First, an all-versus-all blastp analysis is performed using an E-value of 1e-5 as cut off value. Next, for each blastp hit pair a modified blast score is calculated: blast score * cov1 * cov2. Here, cov1 and cov2 are the alignment coverages for protein 1 and 2 of the pair, respectively. This alignment coverage is a fraction of 1. This modified blast score is used as input for the MCL Markov clustering program (Van Dongen, 2000; Enright *et al.*, 2002), using an inflation parameter of 2. Each resulting cluster is considered a multi-gene family. A predicted protein is considered unique to *M. osmundae* when it is a member of a multi-gene family that is only found in that organism. The core proteome was determined by identifying multi-gene families that contained at least one member in each of the analyzed species.

Representation analysis

Custom scripts were developed in Python and R to analyze over- and under-representation of functional annotation terms in sets of genes using the Fisher Exact test. The Benjamini-Hochberg correction was used to correct for multiple testing using a p-value of 0.05, where applicable.

Phylogenetic analysis

As part of the Assembling the Fungal Tree of Life (AFTOL) project, 71 orthologous proteins were identified as single or low copy using a Markov clustering approach (Floudas *et al.*, 2012). Hidden Markov models were constructed for each of the 71 proteins, which were then used to mine the genome of *M. osmundae* for the best-hit protein to each orthologous cluster. These *M. osmundae* proteins were added to an existing dataset (Padamsee *et al.*, 2012) and protein clusters were aligned using MUSCLE (Edgar, 2004). Poorly aligned regions of the alignments were assessed and removed with liberal (maximum number of contiguous non-conserved positions allowed is 8; minimum length of a block allowed is 5) and conservative (maximum number of contiguous non-conserved positions allowed is 4; minimum length of a block allowed is 10) settings of GBlocks (Castresana, 2000). To gauge the accuracy of orthology assessment, RAxML phylogenetic trees were directly examined for each individual protein cluster. Proteins identified as non-orthologous due to well-supported erroneous phylogenetic placement (e.g., outside of Basidiomycota) were removed from subsequent analyses. Liberal and conservative protein alignments were concatenated into a superalignment and analyzed using RAxML with a GAMMAPROT model of evolution. For both liberal and conservative superalignments, an additional PAML analysis was run in which amino acid positions were binned into rate categories 1 through 8 using the codeml option with 8 being the most variable (Yang 2007). Three sets of RAxML analyses were then performed by removing positions from the 8th, 7th and 8th, and 6th, 7th and 8th rate categories to assess the effect on branch support. For all analyses, branch support was assessed using 100 RAxML bootstraps (Stamatakis, 2006).

CAFE analysis

The phylogenetic tree described above was fitted to CAFE's requirements by converting it to an ultrametric tree using TreeEdit v1.0 (<http://tree.bio.ed.ac.uk/software/treededit/>), and converting

the branch lengths to integers using a custom Python script. This tree and the number of proteins having a given Pfam assignment, for a given organism, were used as an input to the program. The search for λ values was run with the $-s$ option, which specifies an optimization algorithm that finds the values of λ that maximize the log likelihood of the data for all families. For all gene families with significant family-wide p-values, CAFE also runs a Viterbi algorithm that assigns branch-specific p-values for expansions and contractions of each family at each branch of the tree.

Mating gene analysis

All four isolates used for mating gene analysis were obtained from the Japan Collection of Microorganisms and maintained on PDA at room temperature. The first three of these represent single-spore isolates from the same collection as the genome specimen (May 1993, Sezawa, Haibara-gun, Nakakawane-cho, Shizuoka Prefecture; elevation 220 m) (Nishida *et al.*, 1995). The fourth culture was isolated two years earlier from the same region but from a different location (Mitsuma) and elevation (200m) (Sugiyama & Katumoto, 2008).

Two gene types that are known to be involved in fungal reproduction were selected for comparison: HD homeodomain genes and the STE3 pheromone receptor genes. Three specific primer sets were designed based on the *M. osmundae* genome sequence: Mo-Sxi2-F (5'-CGGTCGCCTACAAGTCTTTC-3') and Mo-Sxi2-R (5'-TCGTCTTCCTCCTCCATAGC-3') for HD genes; Mo-STE3-F (5'-GCCCTTCTTGTGGCATCTAC-3') and Mo-STE3-R (5'-TGGCAAATTAAACAGCACGA-3') for STE3 gene and MoPh1-F (5'-GCAATGGTCACATTGTGCGAG-3') and MoPh1-R (5'-CACCCCAGGAGAAACAAAAA-3') for MoPh1 gene. These primers were used to amplify the mating type regions from each strain and PCR was conducted using ExTaq DNA polymerase (Takara Bio Inc., Otsu, Shiga Prefecture, Japan) with the following parameters: 94°C for 3 min; 35 cycles of 94°C for 1 min, 56°C for 0.5 min and 72°C for 1 min; elongation at 72°C for 7 min. PCR products were purified using USB ExoSAP-IT (Affymetix, Santa Clara, CA, USA) and sequenced via Sanger sequencing at the University of Michigan Sequencing Core.

Meiosis genes

Meiosis-related genes were determined from previously published sources (Burns *et al.*, 2010; Halary *et al.*, 2011; Schurko & Logsdon, 2008 and references therein), and by searching for the word ‘Meiosis’ in the gene name descriptions on the *Saccharomyces* Genome Database, and the KEGG meiosis pathway obtained from the Kyoto Encyclopedia of Genes and Genomes database (www.kegg.jp). If no match was found for a given protein from *Saccharomyces cerevisiae* Meyen ex E.C. Hansen, an orthologous protein sequence from *Ustilago maydis* (DC.) Corda was used if available (data obtained from Donaldson & Saville, 2008; Güldener *et al.*, 2006).

Supplementary references

- Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S. 2011.** Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7**: e1002230.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* 2000.** Gene ontology: tool for the unification of biology. *Nature Genetics* **25**: 25–29.
- Birney E, Clamp M, Durbin R. 2004.** GeneWise and Genomewise. *Genome Research* **14**: 988–995.
- Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R. 2009.** *Ustilago maydis* as a pathogen. *Annual Review of Phytopathology* **47**: 423–445.
- Caldwell GA, Naider F, Becker JM. 1995.** Fungal lipopeptide mating pheromones: a model system for the study of protein prenylation. *Microbiological reviews* **59**: 406–422.
- Casaregola S, Neuvéglise C, Lépling A, Bon E, Feynerol C, Artiguenave F, Wincker P, Gaillardin C. 2000.** Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*. *FEBS Letters* **1**: 95–100.
- Catanzariti AM, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG. 2006.** Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**: 242–256.
- Dodds PN, Lawrence GJ, Catanzariti AM, Teh T, Wang CIA, Ayliffe MA, Kobe B, Ellis JG.**

2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of National Academy of Sciences of USA* **103**: 8888–8893.

Donaldson ME, Saville BJ. 2008. Bioinformatic identification of *Ustilago maydis* meiosis genes. *Fungal Genetics and Biology* **45**: S47–S53.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575–1584.

Fell JW, Boekhout T, Fonseca A, Sampaio JP. 2001. Basidiomycetous Yeasts. In: McLaughlin DJ, McLaughlin EG, Lemke PA, eds. *The Mycota, Systematics and Evolution*. Springer-Verlag, Berlin, Germany. Vol **7B**: 3–35.

Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Bastürkmen M, Spevak CC, Clutterbuck J *et al.* 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105–1115.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S *et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of National Academy of Sciences of USA* **108**: 1513–1518.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al.* 1996. Life with 6000 genes. *Science* **274**: 546–567.

Güldener U, Münsterkötter M, Oesterheld M, Pagel P, Ruepp A, Mewes H-W, Stümpflen V. 2006. MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Research* **34**: D436–D441.

Hane JK, Lowe RGT, Solomon PS, Tan KC, Schoch CL, Spatafora JW, Crous PW, Kodira C, Birren BW, Galagan JE *et al.* 2007. Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *The Plant Cell* **19**: 3347–3368.

Joneson S, Stajich JE, Shiu SH, Rosenblum EB. 2011. Genomic transition to pathogenicity in chytrid fungi. *PLoS Pathogens* **4**: e1002338.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O & Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**: 462–467.

Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Müller O *et al.* 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*, *Nature* **444**: 97–101.

Kanehisa MGS, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. *Genome Biology* **5**: R7.

Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS *et al.* 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology* **5**: R7.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels* **6**: 41.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**: 1754–1760.

Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA *et al.* 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**: 1321–1324.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**: 955–964.

Ma LJ, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, Elias M, Idnurm A, Lang BF, Sone T. 2009. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genetics* **5**: e1000549.

Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, 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.

Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z. 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* **11**: 663.

Martínez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F *et al.* 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology* **22**: 695–700.

- Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D et al. 2008.** Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology* **26**: 553–560.
- Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P et al. 2009.** Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proceedings of the National Academy of Sciences of USA* **106**: 1954–1959.
- Melen K, Krogh A, von Heijne G. 2003.** Reliability measures for membrane protein topology prediction algorithms. *Journal of Molecular Biology* **327**: 735–744.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne, G. 1997.** Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**: 1–6.
- Ohm RA, Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, de Vries RP, Record E, Levasseur A, Baker SE et al. 2010.** Genome sequence of the model mushroom *Schizophyllum commune*. *Nature Biotechnology* **28**: 957–963.
- Price AL, Jones NC, Pevzner PA. 2005.** De novo identification of repeat families in large genomes. *Bioinformatics* **21**: i351–i358.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005.** InterProScan: protein domains identifier. *Nucleic Acids Research* **33**, W116–W120.
- Ridout CJ, Skamnioti P, Porritt O, Sacristan S, Jones JD, Brown JK. 2006.** Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell* **18**: 2402–2414.
- Salamov AA, Solovyev VV. 2000.** Ab initio gene finding in *Drosophila* genomic DNA. *Genome Research* **10**: 516–522.
- Schurko AM, Logsdon JM Jr. 2008.** Using a meiosis detection toolkit to investigate ancient asexual “scandals” and the evolution of sex. *BioEssays* **30**: 579–589.
- Smit AFA, Hubley R, Green P. 2010.** *RepeatMasker Open-3.0*. 1996–2010 [WWW document] URL www.repeatmasker.org [accessed on 02.03.2013].
- Stajich JE, Wilke SK, Ahrén D, Au CH, Birren BW, Borodovsky M, Burns C, Canbäck B, Casselton LA, Cheng CK et al. 2010.** Insights into evolution of multicellular fungi from the

assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*).

Proceedings of the National Academy of Sciences of USA **26**: 11889–11894.

Ter-Hovhannisyanyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Research* **18**: 1979–1990.

Van Dongen S. 2000. *Graph Clustering by Flow Simulation* PhD dissertation. Utrecht University, Utrecht, The Netherlands.

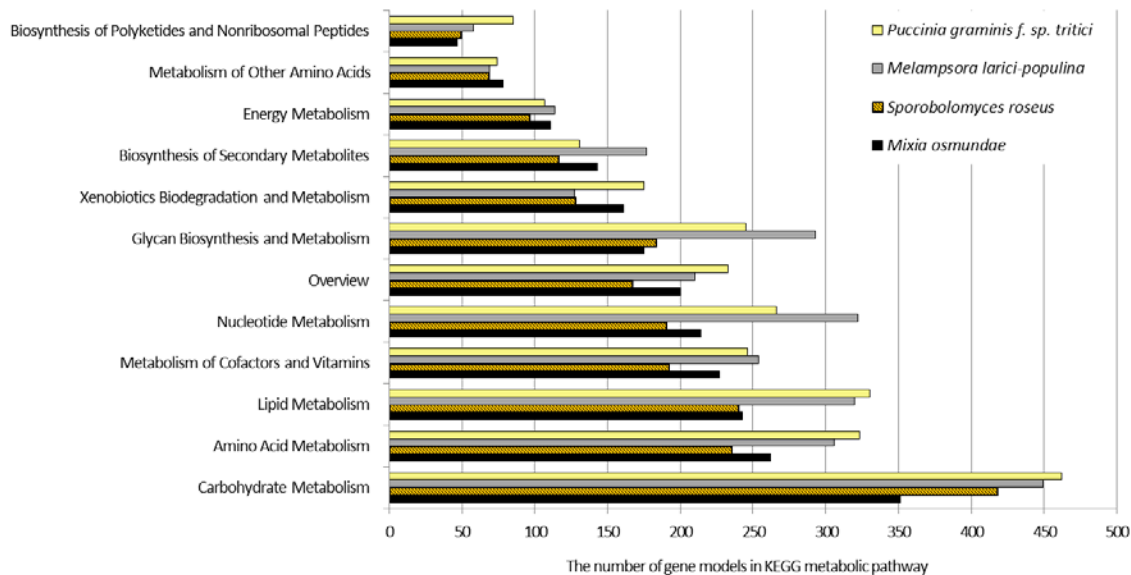
Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S et al. 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871–880.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Zerbino DR, Birney E. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**: 821–829.

Fig. S1 A comparison of KEGG metabolic pathway gene models in Pucciniomycotina. The data presented were obtained from MycoCosm. (a) – number of gene models in each KEGG category; (b) – the % of total number of gene models in each KEGG category.

(a)



(b)

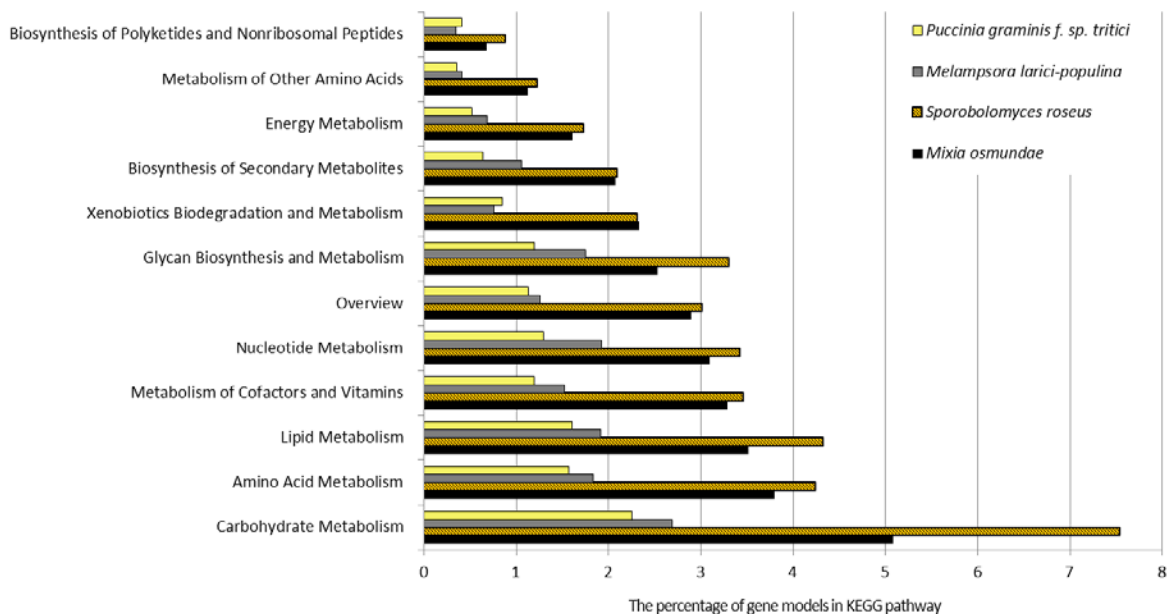


Fig. S2 A comparison of gene models in Pucciniomycotina based on KOG classification. The data were obtained from MycoCosm.

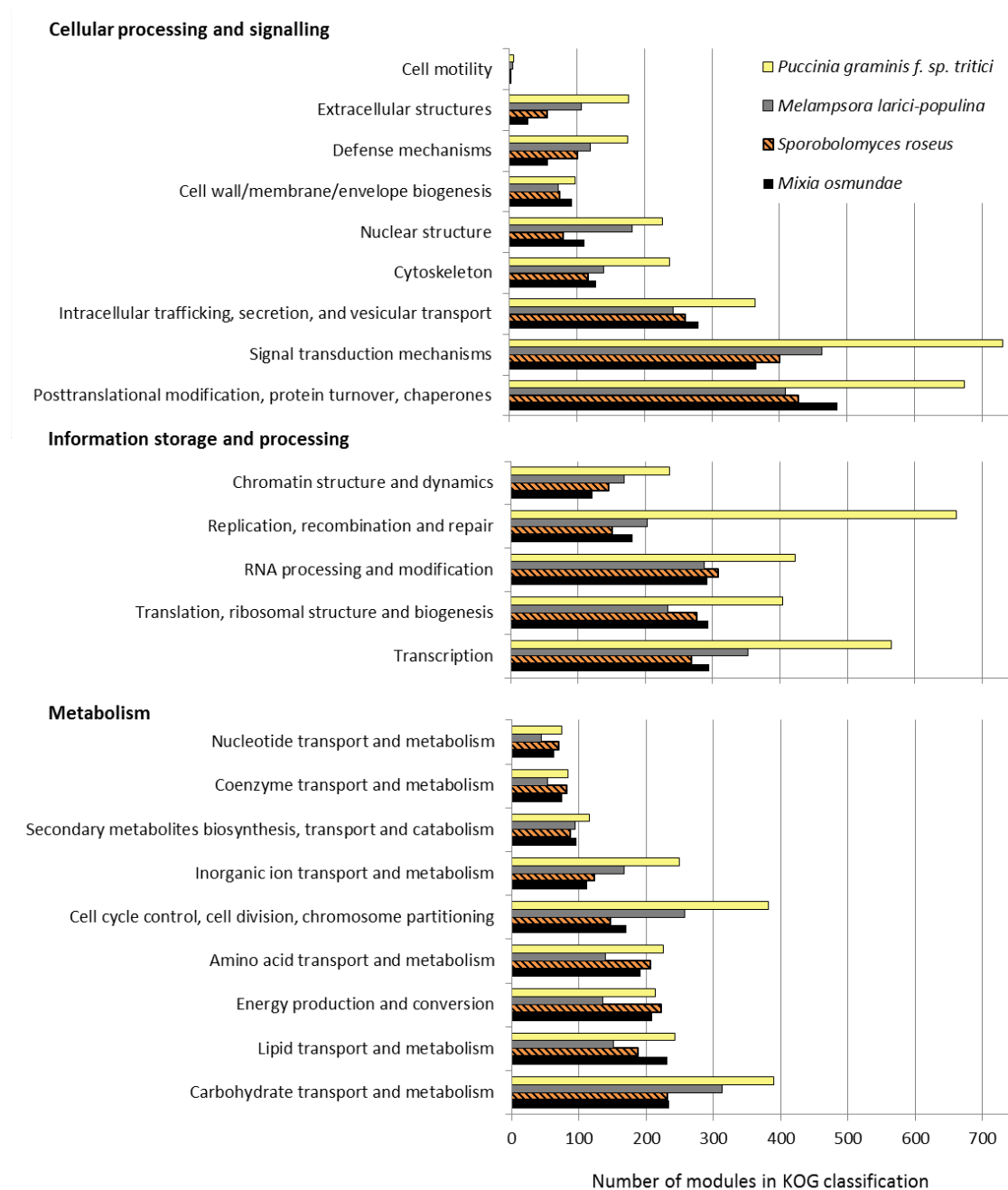


Fig. S3 Core and non-core proteome of the published Pucciniomycotina and Ustilaginomycotina genomes, demonstrating that although *Mixia osmundae* has a small genome, less than 50% of it is comprised of core genes shared with other basal Basidiomycota.

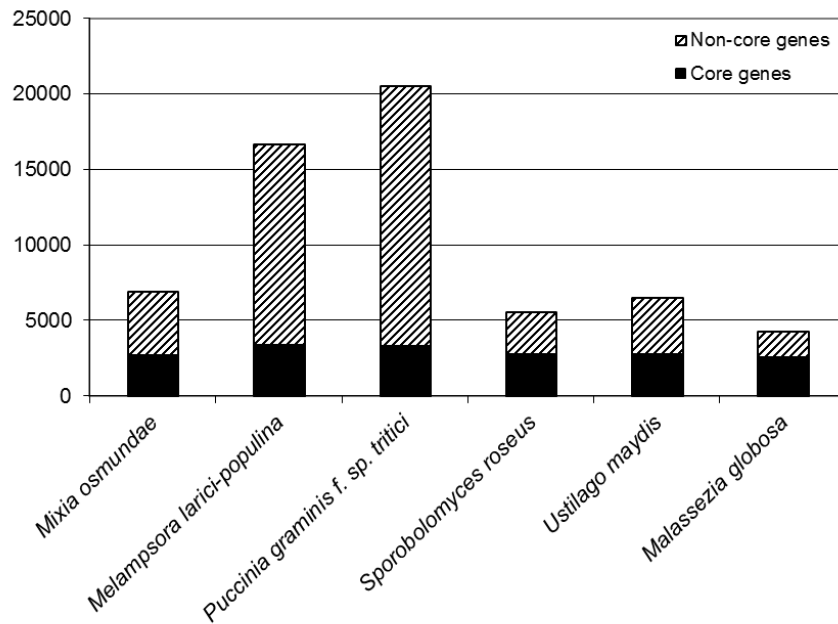


Figure S4 Conservation of *M. osmundae* predicted proteins in related fungi, determined by a multi-gene family analysis. The majority of predicted secreted proteins and SSPs (small secreted proteins) are unique to *M. osmundae*, which contrasts with the conservation of the total proteome. These unique SSPs may be effectors involved in the infection of the *Osmunda* host. Sixteen SSPs are found in at least one other species belonging to the Pucciniomycotina (*S. roseus*, *P. graminis* f. sp. *tritici* and *M. larici-populina*), and 23 are also found in at least one of the Ustilaginomycotina (*U. maydis* and *M. globosa*). Only 10 are part of the core proteome, which means that they are conserved in all six studied Basidiomycota.

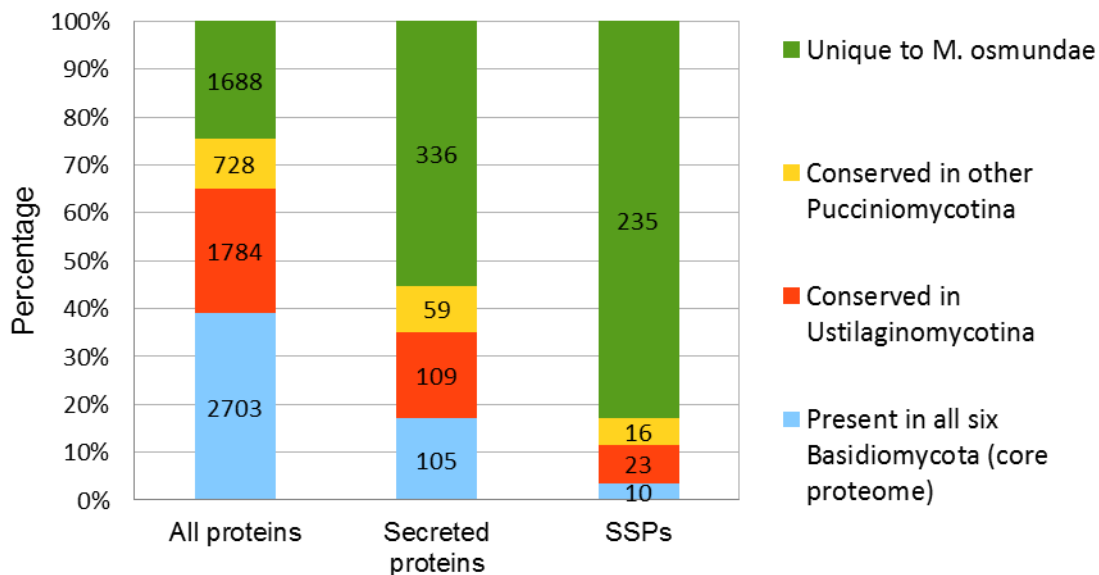


Fig. S5 *Mixia osmundae* pheromone precursors are processed into three mature peptides. Shown are the amino acid sequences of five pheromone precursor proteins: MoPh1&2 = *M. osmundae* pheromone precursors MoPh1 & MoPh2; *Rhodospiridium toruloides* mating type A2 (RHA2.A2); *Melampsora larici-populina* (MlpPh4); *Rhodospiridium toruloides* mating type A1 (RHA1.A1); and *Sporidiobolus salmonicolor* mating type A1 (SsRHA3). Grey regions indicate putative mature pheromones, and the underlined regions are CAAX-motifs for C-terminal processing (Caldwell *et al.*, 1995). Annotations of other proteins follow Kües *et al.* (2011).

MoPh1&2 MSPAPITVEVESVAAPPFAM^{YRGCTIAQTEEVATDARDPPFAM^{YRGCTIAQSEETSDARDPPFAM^{YRGCTIA}}}
RHA2.A2 MAPITLAANQEVAAPALVNEEHNPTVTSYMACTVSRKDVSVEMVDEEHNPTVTSYMACTIA
MlpPh4 MSSATSQKHENGSPSAAIPLNSEREGQVGGGSR^{YCVISQRQGQVGGGSKYCVISQRQGQVGGGSKYCVWS}
Rha1.A1 MVAYPEISWRTNGCTVSK-YPEISWRTNGCTVSK-YPEISWRTNGCTVSK-YPEISWRTNGCTVA
SsRHA3 MVA^{YPTIGVTPGCTVSR}TYPTIGVTPGCTVSKTYPTIGVTPGCTVA

Fig. S6 Comparison of the ITS region of *Mixia osmundae* isolates (sequences 1–4, obtained by direct sequencing) and closest GenBank matches (sequences 5–9). The aligned regions of a portion of ITS1 and ITS2 are shown. The areas not shown were identical between the studied sequences. Colored blocks show the nucleotides that differ between the sequences; framed areas indicate differences between known *M. osmundae* isolates and unidentified environmental isolates from bamboo leaves (5–6) and European beech leaves (7–9). The total length of the alignment was 174 bp for sequences 1–9 and 446 bp for sequences 1–6.

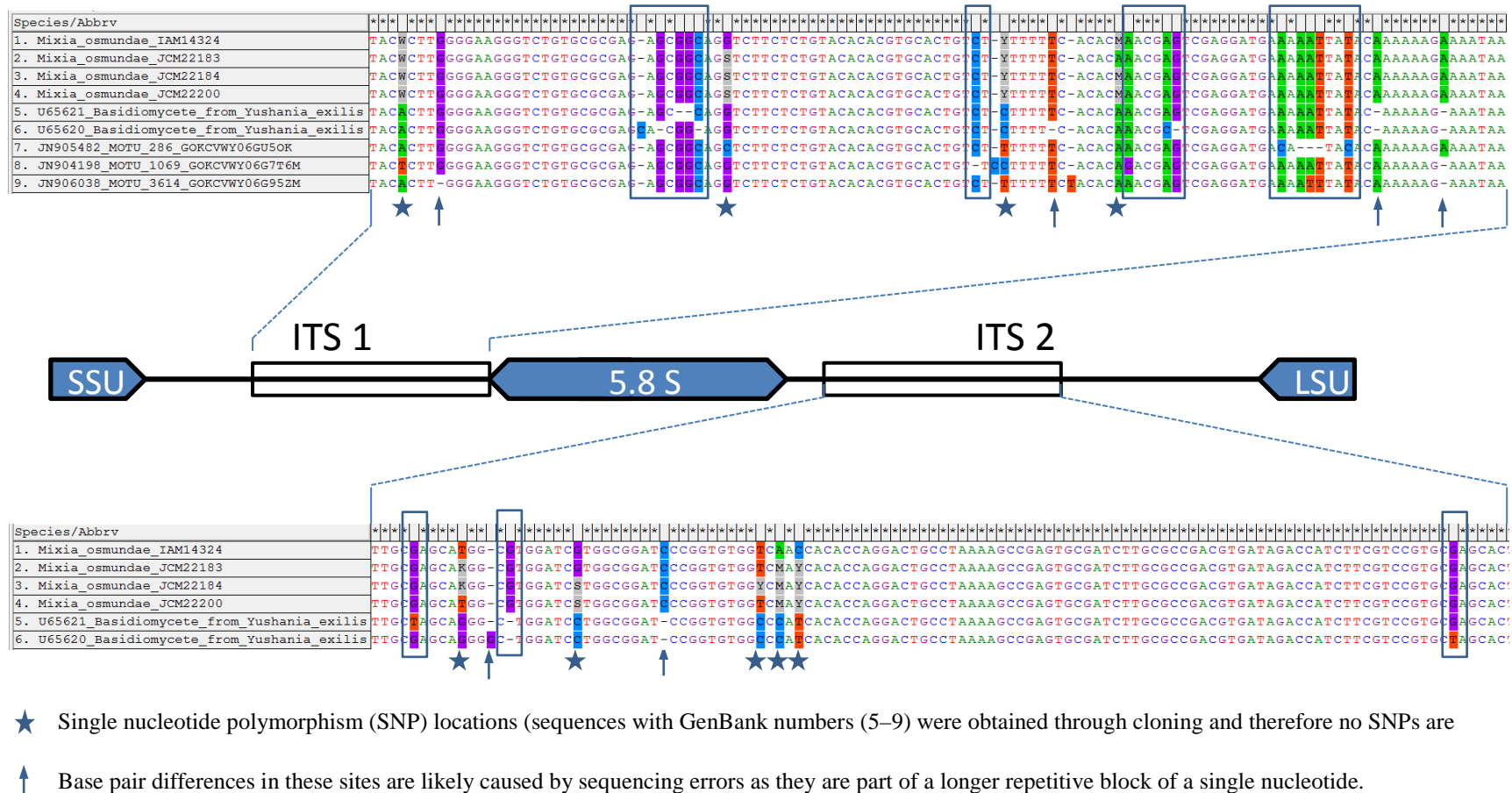


Fig. S7 Meiosis pathway in *Saccharomyces cerevisiae* and the orthologous genes in *Mixia osmundae*. Green squares represent the genes that were also found in *M. osmundae* and pink represent genes that had no orthologs in the sequenced genome. Additional information on search parameters is given in Methods S1. Pathway map was obtained from www.kegg.jp.

