Supporting Information

Stajich et al. 10.1073/pnas.1003391107

SI Text

Genome Assembly. The haploid Okayama 7 #130 strain, a common laboratory strain, was sequenced by whole-genome shotgun (WGS) sequencing of 4-kb and 10-kb plasmids and 40-kb fosmids to a level of 10× coverage of the predicted 36 megabase (Mb) genome. The initial assembly using Arachne (1) was made public July 2003 and consisted of 431 WGS contigs with an N50 of 218,423 kb. The contigs were assembled into 106 scaffolds by mate-pair information. One scaffold comprised the mitochondrial DNA (42,448 nt). Genome finishing was performed by PCR from ends of WGS contigs predicted to be adjacent based on the scaffold assignment and on BAC mapping. There were 157 successful PCR reactions yielding single products spanning two contigs, providing the primary evidence for closing 96 gaps. Re-analysis of the original fosmid reads using Consed (2) and a limited number of additional PCR reactions gave evidence closing 60 gaps between contigs. Alignment of the WGS contigs showed 88 cases in which contigs overlapped. In 91 cases, neighboring contigs were joined in the current assembly by abutting them without a known joining sequence. TERMINUS (3) was used to identify telomeric repeats on contig ends. The chromosomes range in size from 0.98 Mb to 4.15 Mb (Fig. S1), and the genome also contains 1.2 Mb of tandemly repeated rDNA sequences at one end of chromosome VI. Most (9 of 13) chromosomes extend telomere to telomere, three chromosomes are missing one telomere, and one chromosome (chromosome VI) has telomere-linked rDNA sequences at one end and is missing the other telomere (Fig. S1).

The remaining 54 WGS contigs, comprising 337,367 nt or 0.94% of the total sequence, were not placed on chromosomes. These contigs contained transposon sequences (11 contigs) and/or were linked to telomere repeats (12 contigs) using TERMINUS (3). The finished chromosome assembly containing 13 chromosomes and 54 unplaced contigs has an N50 of 3.47 Mb. The initial sequence, assembly, and annotation can be accessed at http://www.broadinstitute.org/annotation/genome/coprinus_cinereus/ . The data in this paper, including annotations, ESTs, serial analysis of gene expression (SAGE) tags, SSRs, linkage data, and oligonucleotides on the microarray platform can be accessed at http://genome.semo.edu/ccin.

BAC Map Construction. A library of BACs was constructed after partial digestion of Okayama 7 #130 DNA with *Hin*dIII and ligation to the pBACTZ vector as described (4). We fingerprinted 1,728 clones using FingerPrinted Contig (FPC) analysis [tolerance 7 and cutoff of 1E-8 (5, 6)]. The FPC output was analyzed using BACFinder (7) which positioned 696 BACs (Dataset S1, Table S1). An additional 303 BACs that could not be placed via FPC were end-sequenced and placed on the scaffolds (Dataset S1, Table S1). The 27 BACs with ends mapping in different scaffolds were used to confirm 17 links between scaffolds (Dataset S1, Table S2). Contigs containing a subtelomeric region were identified using the TERMINUS program (3), which allowed the identification of 21 scaffold ends linked to telomere sequences (Fig. S1 and Dataset S1, Table S2).

Genetic Linkage Map Construction. RepeatMasker (8) was used to identify 537 genomic regions with a minimum of seven perfect trinucleotide repeats or six perfect tetranucleotide repeats. Primers were designed to produce 200- to 250-nt fragments flanking 404 of these simple sequence repeats (SSRs) and were used to amplify DNAs including the SSRs from Okayama 7 #130 (the sequenced strain, *A43B43 ade3*) and HT 14.01 #172 (the mapping partner, *A6B6 trp1*). Polymorphisms ranging in size from 3 to 30 nt were identified for 133 regions, which were used as markers to construct the linkage map. Forward primers with Big Dye fluorophores added to the 5′ end were used to amplify DNAs from the two parental strains and 46 random spore progeny in 96-well format. PCR reactions (20 μL total) contained 8 ng genomic DNA, 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 2 mM MgCl2, 0.25 μM sense primer, 0.25 μM antisense primer, 0.8 mM dNTPs, and 0.1 U Taq polymerase (Invitrogen). Cycling conditions were 95 °C for 3 min followed by 10 cycles of 94 °C (30 s), 62–52 °C (30 s; 1° decline each cycle), 72 °C (45 s), then followed by 29 cycles of 94 °C (30 s), 52 °C (30 s), 72 °C (45 s) followed by a 20-min hold at 72 °C. Fragments were resolved using an ABI 3730XL Genetic Analyzer and scored using GeneMapper software (Applied Biosystems). In addition to the polymorphic SSRs, we included the two mating-type loci (*A* and *B*) and two nutritional markers (*ade3* and *trp1*) in the map. The alleles at the mating-type loci were determined by crossing each random spore to four mating-type tester strains (*A43B43*, *A6B6*, *A43B6*, *A6B43*) and observing clamp connections (Fig. 3*D*) in the compatible mating (e.g., a segregant with the *A43B6* mating type forms clamp connections only with the *A6B43* tester). The alleles at *ade3* and *trp1* were determined by growth on minimal medium, minimal medium plus adenine, minimal media plus tryptophan, and minimal medium plus both adenine and tryptophan.

Segregation patterns for all the markers were examined, and none exhibited segregation distortion. We used MAPMAKER software (9) to construct the linkage map. The “group” command with a minimum LODlinkage of 3.0 and maximum distance of 60 centimorgans (cM) produced 13 linkage groups that corresponded to the 13 chromosomes, with two exceptions. First, we observed pseudolinkage of the right arm of chromosome IV and chromosome XIII. Contour-clamped homogenous field electrophoresis revealed that in strain HT14.01 #172, chromosome IV is smaller and chromosome XIII is larger than their corresponding chromosomes in strain Okayama 7 #130, consistent with a terminal translocation. Four segregants exhibited two alleles for the terminal SSR on chromosome IV, indicating that they inherited a normal chromosome and a translocated chromosome. Second, the markers on chromosome VIII fell into two groups. Because linkage between these two groups was well supported by the sequence assembly, we assumed a map distance of 50 cM for this interval. The chromosome positions of the SSRs and other markers, map distances, and calculated kb/cM are shown in Dataset S1, Table S6. Although a higher marker density in some genomic intervals might reveal additional hotspots (if crossover distribution is non-uniform in these intervals), our results are in good agreement with the linkage map produced using 219 random amplified polymorphic DNA (RAPD) markers (10). Every chromosome in the RAPD map displayed an internal cluster of RAPD markers (40% of all markers) that failed to recombine, as predicted if the RAPD markers are randomly distributed in the *Coprinopsis cinerea* genome and if 45% of the genome undergoes very little meiotic recombination. The linkage map was supported further by the analysis of four tetrads which placed 60 crossovers. Only four of these crossovers occurred in cold regions, whereas 22 occurred in hot regions. The linkage map also supported 11 links between scaffolds (Dataset S1, Table S2).

Transposon Identification. Retrotransposons (class I transposable elements) were identified using PILER (11), RepeatScout (12), and RepeatRunner (13). PILER generated a library of 29 repeat families. BLASTX (14) analysis of those families showed that five families showed strong hits to retrotransposons, five families showed weak hits to various proteins, and 19 families showed no significant hits. Four of the five families (0.27, 0.28, 23.4, and 6.21) showed similarity to *Gypsy* elements, whereas one family (10.17) showed similarity to *Copia* elements. Twelve additional *Copia* elements were found using a keyword search of the *C. cinerea* annotations. RepeatMasker (8) was used to determine the copy number and chromosome distribution of these families (Dataset S1, Tables S3a and S3b). A total of 834 kb is found in the five retrotransposon families. No DNA transposons (class II transposable elements) were identified through the previous analysis. A search of Pfam domains and gene predictions of the *C. cinerea* annotations was conducted to identify any potential DNA transposons. Initially, 26 potential elements were identified by this search method. The elements fall into three classes of DNA transposons: Activator (hAT), Enhancer (En/spm), and Mariner (Tc1). A total of 70 kb is found in the three DNA transposon families. The copy number and chromosome distribution of these families is shown in Dataset S1, Tables S3a and S3b.

Genome Annotation. The protein-coding genes (13,342 sequences) were identified using a combination of gene prediction and evidence-based tools. The v1 annotations generated required extensive training because of limited curated gene models for *Coprinopsis* or any Agaricomycete fungus at the time (2002). The initial gene set was produced on the first version of the assembly (NZ_AACS00000000.1) and used ab initio predictors SNAP (15), AUGUSTUS (16), Twinscan (17), and GeneZilla (18), which were trained using a set of gene models predicted from protein-to-genome alignments generated by Genewise (19). Twinscan models were predicted with the basidiomycete *Phanerochaete chrysosporium* as an informant genome. The comparative, computational, and experimental data sets were combined into a final gene call using the tool GLEAN (20) that uses a latent classification scheme to scale the contribution of each type of evidence and to assemble the exons into the largest ORF. We constrained GLEAN to limit the size of the introns to 300 nt.

Additional gene calling was performed on the second version of the assembly (AACS02000000) with the Broad Institute gene-calling pipeline using additional EST data, GeneMark.hmm+ES version 2 (21), FGENESH (22), and GENEID (23). Previously predicted genes were used to identify genes that may have been missed and potential splits and merges. Mitochondrial genes were predicted using GeneWise (19) and computational or manual identification of ORFs based on BLAST (24) and Pfam homology (25) to reveal mitochondrial genes. Repeat regions were identified based on multiple alignments of transcripts to the whole genome and overlap with BLAST and Pfam hits from transposon proteins or domains. Genes flagged by these screens were excluded from the final gene set based on manually determined cutoff points. Predicted proteins without supporting evidence were excluded from the final gene set if their coding sequence (CDS)/transcript length ratio was less than 0.33 or their CDS length was less than 80 residues.

The 267 tRNA genes were identified using tRNAScan-SE. These genes are distributed randomly on all of the chromosomes and show no tendency to be colocalized with transposable elements. Nuclear and mitochondrial ribosomal RNA genes were identified based on comparison with highly similar ribosomal RNA sequences from *Coniophora puteana* (GenBank accession AM946631), *Agrocybe aegerita* (AAU54637), and *Suillus sinuspaulianus* (L47585), supplemented with predictions from RNAmmer (26) and Rfam (27).

ESTs. Transcript data from 5,612 ESTs were employed to support gene calls. We used nine different growth conditions for our EST libraries. Libraries CCFBM and CCK+6 were obtained from a dikaryon constructed from strains backcrossed to Okayama7 #130 for five generations (J6, 5–4 #409 × J6, 5–5 #410). The dikaryon was grown on yeast extract/malt extract/glucose (YMG) medium (28) at 37 °C until confluent and then was transferred to 25 °C with a 16-h light/8-h dark cycle.

CCFBM: Fruit body caps were harvested either 1 h before or 1 h after karyogamy. Poly-A^+ RNA was isolated from both groups and pooled to make cDNA, which was cloned into pBluescript II SK- phagemids. There are 1,175 accessions in the GenBank EST database from library CCFBM (DR774668–DR775517).

CCK+6: Fruit body caps were harvested 6 h after karyogamy. Library construction was performed as for CCFBM. There are 1,667 accessions in the GenBank EST database from library CCK+6 (DN591505–DN593171).

All other libraries were constructed using *C. cinerea* strain Okayama 7 #130 monokaryotic mycelia and, except as noted, were cultured at 37 °C for 3 d on minimal medium before harvest or additional treatment. Harvested mycelia were frozen in liquid nitrogen and stored at −80 °C before RNA extraction. RNA was extracted using the Qiagen RNeasy plant mini kit per kit instructions. Libraries were constructed using the Stratagene Bluescript II XR cDNA library construction kit per kit instructions. To minimize repeated sequencing of clones from highly expressed mRNAs in these libraries, we chose 47 predicted genes that had the highest number of ESTs in the CCFBM, CCK+6, and CCMIN libraries (which were not subtracted). Oligonucleotide probes were designed for these genes (which included hydrophobins, ribosomal proteins, and many other "housekeeping" genes) and were used to screen the remaining libraries. This method reduced the redundant sequencing from 10% in the initial unsubtracted libraries to 2% in the subtracted libraries.

CCMIN: Mycelia were cultured at 37 °C for 3 d on minimal medium (29) before harvest. There are 670 accessions in the GenBank EST database from library CCMIN (DN593172–DN593841).

CCYMG: Mycelia were cultured at 37 °C for 3 d on YMG medium before harvest. There are 550 accessions in the GenBank EST database from library CCYMG (DR752715–DR753264).

CCRAP (rapamycin): Mycelia were cultured at 37 °C for 3 d on minimal medium and then were transferred to minimal medium with 100 mM rapamycin for 1 h before harvest. There are 747 accessions in the GenBank EST database from library CCRAP (DN593842–DN593917, DR753265–DR753301, and DR753303–DR753936).

CCOS (osmotic shock): Mycelia were cultured at 37 °C for 3 d on minimal medium and then were transferred to minimal medium with 1M sorbitol for 1 h before harvest. There are 62 accessions in the GenBank EST database from library CCOS (FG068230–FG068291).

CCHS (heat shock): Mycelia were cultured at 37 °C for 3 d on minimal medium and then were transferred to prewarmed minimal medium and incubated at 42 °C for 1 h before harvest. There are 540 accessions in the GenBank EST database from library CCHS (DR421062–DR421601).

CCCN (complex carbon/nitrogen source): Mycelia were cultured at 37 °C for 3 d on minimal medium with 2% wt/vol cellobiose and 0.4% wt/vol gelatin substituting for glucose and L-asparagine. There are 560 accessions in the GenBank EST database from library CCCN (DR752151–DR752714 and DR753937–DR753939).

CCSEN (senescent): Mycelia were cultured at 37 °C on minimal medium. After 4 d, the extent of the mycelium was marked on the plate. After 9 d, the first 4-d growth was harvested. There are 560 accessions in the GenBank EST database from library CCSEN (DR907568–DR908072).

SAGE Library Construction. To confirm the gene calls and to examine the regulatory complexity in *C. cinerea*, we prepared 5′ SAGE libraries from vegetative (dikaryotic) tissue and from fruit body primordia. A dikaryon using strains backcrossed to Okayama 7 #130

(J6, 5–4 #409 × J6, 5–5 #410) was cultivated on YMG medium (28). The mycelium was cultured on agar plates at 37 °C for about 7 d until the mycelium covered the whole agar surface. The primordium was induced by incubating the mycelial culture at 25 °C under a light/dark regime of 14/10 h. The incubator was kept at a relative humidity higher than 60%. Total RNA was extracted from mycelia when they grew over the whole agar surface and from stage 1 primordia when they grew to a height of about 5 mm. The RNAs were extracted from tissue frozen in liquid nitrogen using TRI reagent (Molecular Research Center, Inc.) followed by chloroform extraction and precipitation. SAGE library construction and analysis were as described (30), with some modifications. Poly-A^+ RNA was isolated using the PolyATract mRNA isolation system (Promega) following the manufacturer's protocol. First-strand cDNA was synthesized using SuperScript III First-Strand Synthesis System for qRT-PCR (Invitrogen). Two separate first-strand synthesis and template-switching (TS) reactions were applied for each developmental stage (mycelium and primordium). The TS oligos were A: 5′-GGGATTTGCTGGTGCAGTACAGGATCCGACggg-3′; B: 5′-GCTGCTCGAATTCAAGCTTCTGGATCCGACggg-3′, where 'g' stands for ribonucleotide. Second-strand cDNA synthesis was performed by low-cycle primer extension using Advantage 2 polymerase (Clontech). Oligos used were CDS primer: 5′-CAGTGGTATCAACGCAGAGTAC(dT)20VN-3′, Anchor primer A: 5′-GGGATTTGCTGGTGCAGTACAGGATCCGAC-3′; Anchor primer B: 5′-GCTGCTCGAATTCAAGCTTCTGGATCCGAC-3′. PCR cycling conditions were 72 °C for 5 min; 95 °C for 45 s; 95 °C for 10 s, 55 °C for 30 s, 68 °C for 4 min for five cycles and 68 °C for 3 min. The PCR products then were purified using QIAquick PCR purification kit (Qiagen). Samples were digested with *Mme*I (New England BioLabs), and 50-bp bands were recovered after acrylamide gel electrophoresis. The cDNAs were ligated to from 100-bp ditags and were amplified using anchor primers A and B with the following cycling conditions: 95 °C for 2 min; 95 °C for 30 s, 65 °C for 45 s, 72 °C for 20 s for 10 cycles, and 72 °C for 3 min. The ditags were purified by phenol/chloroform extraction and ethanol precipitation and were used for high-throughput pyrosequencing (454 Life Sciences GS20 sequencer). Individual tags were extracted from the ditag and checked for sequence quality (Phred equivalent >20), and the unmapped starting G residues were removed. The resulting tags were aligned to the genome. Predicted genes with two or more tags <500 nt upstream of the start codon are recorded as "sense" in Dataset S1, Table S5; predicted genes with two or more tags <500 nt downstream of the stop codon or within the coding region are recorded as "antisense" in Dataset S1, Table S5. Strand-specific PCR confirmed the presence of four predicted antisense transcripts.

Gene Families. A similarity search of all protein-coding genes in *C. cinerea*, *Laccaria bicolor* (31), and *P. chrysosporium* (32, 33) was performed using National Center for Biotechnology Information (NCBI) BLASTP (24) with an E-value cutoff of $1E^{-10}$ as input for the clustering of proteins into protein families. TribeMCL (34) was run with default inflation value (1.5) and generated 7,433 protein families (at least two members in a family) and 5,044 singletons from *C. cinerea*, *L. bicolor*, and *P. chrysosporium*. The functional annotation using Pfam showed that families with hydrophobin domain and P450 domain are among the most expanded in *C. cinerea* (Dataset S1, Table S8).

Protein Kinases. A preliminary set of *C. cinerea* protein kinases was delineated using hmmsearch (version 2.3.2) (35, 36) to screen the predicted expressed proteins with a hidden Markov model (HMM) derived from an alignment of the complete *Dictyostelium discoideum* kinome (37) using an E-value cutoff of 1. Additional protein kinases with E-values between $1E^{-1}$ and $1E^{-100}$ were identified based on the conservation of protein kinase subdomains (38) determined by manual inspection of the alignment to the HMM alignment. A six-frame translation of the genome also was screened with a library of atypical and divergent protein kinases (kin20.hmm, available at www.kinase.com). Translated protein kinases and their classifications, locations, and correspondence with the previous annotation release are presented in Dataset S1, Table S9a. Genes that were missed in initial annotation are denoted NewKin with a numeric designation.

Kinases were classified based on BLAST (24), and class-specific HMMs; ambiguous cases were resolved based on differential BLAST/HMM scores and manual examination of multiple sequence alignments. Novel *C. cinerea* kinases were clustered using OrthoMCL (39) to form seed groups, which were aligned and used to build custom HMMs. The *C. cinerea* genome, the nonredundant protein sequence database at NCBI, and fungal databases were screened against these custom HMMs to identify novel *C. cinerea* kinases missed in the initial screens, and to identify novel kinases in other species. Subfamilies within the FunK1 family were delineated using a neighbor-joining tree (40). The numbers of family and subfamily members in *C. cinerea* are compared with those from a diverse set of species with complete or draft kinomes in Dataset S1, Table S9b. Differences between the *C. cinerea* kinome and the extensively characterized *Saccharomyces cerevisiae* kinome are highlighted in Dataset S1, Table S9c.

The FunK1 family members are of particular interest, because these kinases appear to be restricted to multicellular fungi. Differential transcription of several of these family members has been noted during specific steps of dikaryon formation (main text and Dataset S1, Table S11). To date, differential transcription of the two FunK1 family members that lie within the *B* locus (Fig. S6) has not been detected. A comparison of FunK1 and conventional kinases was performed to examine whether key regulatory residues were conserved (Fig. 3). The HMMlogo was constructed using HMMeditor (41) using structure 1ATP (42) numbering. *C. cinerea* has 16 members of the tyrosine kinase-like (TKL) kinase group, which are the likely progenitor of animal tyrosine kinases. This group of kinases is found in the Basidiomycetes but is entirely absent from the Saccharomycotina. In plants and *D. discoideum*, TKLs often function as receptor kinases, but no fungal members appear to be membrane linked. In fact, no conventional or atypical *C. cinerea* kinase is predicted to contain a transmembrane helix domain; if any of these kinases transduce intercellular signals, they presumably do so through a distinct set of nonkinase receptors.

L. bicolor protein kinases were manually curated previously (43), and the total counts per family are presented in Dataset S1, Table S9b.

A phylogenetic tree of MAPK also was constructed to check for the presence of *FUS3* and *HOG1* orthologs in *C. cinerea* (Fig. S3). Examination of the tree revealed that orthologs of both are found in *C. cinerea* and that several duplications of the *FUS3* family have occurred in the Basidiomycetes.

Phylogenetic Analysis of P450 and Hydrophobin Genes. The gene family expansions from the TRIBE-MCL clusters showed obvious expansions of the number of P450 domain-containing genes in *C. cinerea*. A large count of P450 family members has been observed previously in *P. chrysosporium* (44), but the extent to which the expansion in the Agaricales is independent has not been explored. To identify the total number of genes containing the P450 Pfam domain, genes were counted using hmmsearch and the P450 HMM profile. The Pfam count did not exactly match the number of *C. cinerea* genes in TRIBE-MCL clusters labeled P450, presumably because of some extraneous sequence additions, but the trend of large numbers is found by both methods. The hmmsearch identifies the full complement of P450 members, whereas the TRIBE-MCL approach can help identify subfamilies. For example, Cluster 2 is a large family in both *C. cinerea* (70 copies) and *P. chrysosporium* (62 copies) but is a smaller family in *L. bicolor* (27 copies).

To examine the physical clustering of genes in the same family, we computed the number of adjacent genes in each of the P450-containing TRIBE families. Cluster 2 has 70 total members in *C. cinerea*, with six pairs of immediately adjacent genes on chromosome II, three pairs on chromosome IX, two pairs on chromosome VI, and one pair on chromosome IV. Cluster 75 has 14 total genes from *C. cinerea,* with two sets of four adjacent genes on chromosome X (CC1G_01582–CC1G_01585; CC1G_01619–CC1G_01622). There are additional examples of adjacent gene pairs for cluster 114 with three separate clusters containing seven total genes on chromosome VIII and a pair on chromosome V for cluster 31.

To study the evolutionary history of the families, we used MrBayes (45) to construct phylogenetic trees of many of the subfamilies of P450 as defined by TRIBE-MCL clusters (Fig. S4). The phylogenetic tree from cluster 75 shows the pattern of species-specific duplications where most of the members coalesce within a species, indicating the gene family diversified after divergence from the ancestor of *P. chrysosporium* and *C. cinerea*. Furthermore, the duplication has occurred independently in both the *P. chrysosporium* and *C. cinerea* lineages, because both species show the pattern of gene family radiations in separate clades. Many members of the families are found near each other on the chromosome in tandem arrays, indicating that in some cases local gene duplication drives the family expansion.

A similar analysis was explored in the hydrophobin gene family. These genes also showed patterns of genomic adjacency, indicating that local gene duplication drives the expansion of the hydrophobin family as well. Comparison of the copy number of hydrophobins across the Basidiomycetes also showed a large expansion in the Agaricomycetes fungi sampled relative to *Ustilago maydis* (one copy) (46) or *Cryptococcus neoformans* (no copies) (47). Family 29 contains most of the hydrophobins, with a total of 34 members from *C. cinerea* but only 17 and 15 for *L. bicolor* and *P. chrysosporium*, respectively. A phylogenetic tree of the relationships of these sequences can be seen in Fig. S5 showing that for the most part there are well-supported clades of species-specific gene duplications indicating independent duplication and expansion of the family in each of the lineages represented by these genomes. Starred genes indicate groups of immediately adjacent genes. The group marked with stars indicates two sets of adjacent genes that are found on chromosome X; one is made up of five genes from CC1G_02181–CC1G_02185, and nearby is a second pair of genes, CC1G_02174 (marked with an open star) and CC1G_02173. The close proximity of the two sets of genes indicates there was a hot spot for duplication of hydrophobins between 1.84–1.86 Mb on chromosome X.

Gene Ontology Analysis of Genes in Regions with Different Recombination Rates. We examined genes within 11 genomic regions with low rates of genomic recombination (> 70 kb/cM, vs. the average 198 kb/cM) ranging in size from 0.8 to 2.4 Mb with 271–996 genes/region (Dataset S1, Table S6). We also examined the 737 genes found in genomic regions with high rates of genomic recombination (< 10 kb/cM, vs. the average 6 kb/cM). These 11 "pseudoclusters" were examined using EASE analysis (48) running in the MeV platform (49) to determine potential enrichment of gene ontology (GO) classes for the 11 sets of linked genes and the set of genes found in regions with high rates of genetic exchange.

It has long been recognized that a reduced recombination rate would be advantageous when particular combinations of alleles at different loci provide a selective advantage and/or other combinations present a disadvantage (50–52). Evidence that genes under such "epistatic selection" are in regions that are cold for recombination also has been noted in the yeast *S. cerevisiae* (53, 54). We tested if genes involved in particular biological processes, cellular components, or molecular functions were found in specific genomic regions with low recombination and if the associated GO terms were distinct from GO terms found associated with genes in freely recombining regions. The cold regions were enriched ($P <$ 0.000005 to $P < 0.01$) for a distinctive combination of functions annotated to basic cellular processes such as helicase activity, Endoplasmic Reticulum-to-Golgi transport, RNA processing, or nitrogen metabolism (Dataset S1, Table S10). In contrast, genes annotated to other processes such as defense responses and cell wall catabolic processes were enriched in the regions of the chromosomes that exhibit high levels of meiotic recombination (Dataset S1, Table S10).

Dikaryon Formation (Mating) and Microarray Analysis. Mating compatibility in *C. cinerea* involves a complex program that is controlled by two sets of unlinked genes, *A* and *B*. To identify downstream targets of these factors, we examined transcripts expressed in mycelia in which the *A*- or *B*-controlled parts of the pathway are activated separately. Strain Okayama 7 #424 (*A43 B43*) was crossed to strain #425 [*A*mut*B*mut (55)], and strains #422 (*A*mut*B43*) and #423 (*A43B*mut) were recovered. The four strains were cultured on YMG, and RNA was isolated as described (EST libraries). First-strand cDNA synthesis, Alexa-Fluor labeling using the Superscript Indirect cDNA Labeling System (Invitrogen), and array hybridization were performed. The two-channel hybridizations compared a sample (*A*mut*B43*, *A43B*mut, or *A*mut*B*mut) with the *A43B43* reference, and four replicates for each sample were analyzed. Data capture and analysis of the 12 arrays was performed with GenePix 4200A scanner (Molecular Devices) and identified with GenePix Pro software (Molecular Devices) at Indiana University Center for Genomics and Bioinformatics. Spots were flagged for omission using GenePix software if they were scored as manually flagged, spots not found, sum of medians <200, or spot pixels <40. Data for a given oligonucleotide were included if two or more of the four replicates contained data for both probes. Of the 13,230 array probes, 11,726 fulfilled these criteria for the *A*mut*B43* arrays, 11,798 for the *A43B*mut arrays, and 10,055 for the *A*mut*B*mut arrays. Significance analysis of microarrays (56) was used to determine significant differences in gene expression between the sample and the control. Expression ratios were $\log_2$ transformed, and Dataset S1, Table S11 reports genes with median (sample/reference) $\log_2 >2$ [false-discovery rate (FDR) < 2%] in six categories (up-regulated in *A*mut*B43*, down-regulated in *A*mut*B43*, up-regulated in *A43B*mut, down-regulated in *A43B*mut, up-regulated in *A*mut*B*mut, and down-regulated in *A*mut*B*mut, with respect to the *A43B43* reference in each case). As expected, *pcc1* and *clp1* were up-regulated in the *A*mut*B43* category, although *clp1* was significant at a FDR of 5% and thus does not appear in Dataset S1, Table S11. None of these categories included three or more genes in tandem, except for the two cases reported in the text.

Divergence Time Estimation. The divergence time estimate of 200 My between *L. bicolor* and *C. cinerea* was estimated using the application r8s (57) and a pruned version of the previously published fungal tree of life (58) with key basidiomycete lineages retained. The maximum age of the Basidiomycota was set to 550 Mya based on results from Taylor and Berbee (59), and the Penalized Likelihood model was applied. The resulting estimate for the divergence between *C. cinerea* and *L. bicolor* is 194 Mya, for the origin of the Agaricomycetes (*C. cinerea*, *L. bicolor*, and *P. chrysosporium* common ancestor) is 317 Mya, and for origin of the Agaricomycotina (Agaricomycetes and *C. neoformans* common ancestor) is 515 Mya. *U. maydis* and *Sporobolomyces roseus* were included as additional outgroups in the Basidiomycetes. These time estimates are still useful comparison points for relative ages of fungi despite suffering from too little fossil data and incomplete sampling of lineages in the Basidiomycetes and outgroups. More detailed studies of fungal divergence times will help to establish robust date estimates.

Synteny. Syntenic regions were identified between *C. cinerea* and *L. bicolor* using FISH (for "Fast Identification of Segmental Homology") (60) based on BLASTP searches with a cutoff of $1E^{-5}$. FISH was run with default parameters, except that we required the minimal block to contain at least four anchors (Dataset S1, Table S12a). Initially, we observed 14 blocks with more than 15 anchors in each (Dataset S1, Table S12b). However, in four cases (S1.1, S2.1, S5.2, and S5.3), two blocks with more than 15 anchors were nearly adjacent in the genome and were treated as a single block for the GO analysis. GO analysis was as described above.

1. Batzoglou S, et al. (2002) ARACHNE: A whole-genome shotgun assembler. *Genome Res* 12:177–189.
2. Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. *Genome Res* 8:195–202.
3. Li W, Rehmeyer CJ, Staben C, Farman ML (2005) TERMINUS—Telomeric End-Read Mining IN Unassembled Sequences. *Bioinformatics* 21:1695–1698.
4. Muraguchi H, Kamada T, Yanagi SO (2005) Construction of a bacterial artificial chromosome (BAC) library of *Coprinus cinereus*. *Mycoscience* 46:49–53.
5. Marra MA, et al. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res* 7:1072–1084.
6. Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res* 10:1772–1787.
7. Crowe ML, Rana D, Fraser F, Bancroft I, Trick M (2002) BACFinder: Genomic localisation of large insert genomic clones based on restriction fingerprinting. *Nucleic Acids Res* 30:e118.
8. Smith AFA, Hubley R, Green P . RepeatMasker Open-3.0. 1996–2010. http://repeatmasker.org.
9. Lander ES, et al. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181.
10. Muraguchi H, Ito Y, Kamada T, Yanagi SO (2003) A linkage map of the basidiomycete *Coprinus cinereus* based on random amplified polymorphic DNAs and restriction fragment length polymorphisms. *Fungal Genet Biol* 40:93–102.
11. Edgar RC, Myers EW (2005) PILER: Identification and classification of genomic repeats. *Bioinformatics* 21(Suppl 1):i152–i158.
12. Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
13. Smith CD, et al. (2007) Improved repeat identification and masking in Dipterans. *Gene* 389:1–9.
14. Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Nat Genet* 3:266–272.
15. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
16. Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19(Suppl 2):ii215–ii225.
17. Korf I, Flicek P, Duan D, Brent MR (2001) Integrating genomic homology into gene structure prediction. *Bioinformatics* 17(Suppl 1):S140–S148.
18. Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20:2878–2879.
19. Birney E, Clamp M, Durbin R (2004) GeneWise and genomewise. *Genome Res* 14: 988–995.
20. Elsik CG, et al. (2007) Creating a honey bee consensus gene set. *Genome Biol* 8:R13.
21. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* 18:1979–1990.
22. Solovyev V, Salamov A (1997) The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proc Int Conf Intell Syst Mol Biol* 5:294–302.
23. Guigó R, Knudsen S, Drake N, Smith T (1992) Prediction of gene structure. *J Mol Biol* 226:141–157.
24. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
25. Finn RD, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36 (Database issue):D281–D288.
26. Lagesen K, et al. (2007) RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108.
27. Gardner PP, et al. (2009) Rfam: Updates to the RNA families database. *Nucleic Acids Res* 37 (Database issue):D136–D140.
28. Rao PS, Niederpruem DJ (1969) Carbohydrate metabolism during morphogenesis of *Coprinus lagopus* (sensu Buller). *J Bacteriol* 100:1222–1228.
29. Moore D, Pukkila PJ (1985) *Coprinus cinereus*: An ideal organism for studies of genetics and developmental biology. *J Biol Educ* 19:31–40.
30. Zhang Z, Dietrich FS (2005) Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5′ SAGE. *Nucleic Acids Res* 33:2838–2851.
31. Martin F, et al. (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452:88–92.
32. Martinez D, et al. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* 22:695–700.
33. Vanden Wymelenberg A, et al. (2006) Computational analysis of the *Phanerochaete chrysosporium* v2.0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins. *Fungal Genet Biol* 43:343–356.
34. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584.
35. Eddy SR. HMMER, version 2.3.2. 1995-2010. http://hmmer.janelia.org/.
36. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
37. Goldberg JM, et al. (2006) The *Dictyostelium* kinome—Analysis of the protein kinases from a simple model organism. *PLoS Genet* 2:e38.
38. Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J* 9:576–596.
39. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
40. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
41. Dai J, Cheng J (2008) HMMEditor: A visual editing tool for profile hidden Markov model. *BMC Genomics* 9(Suppl 1):S8.
42. Zheng J, et al. (1993) 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr D Biol Crystallogr* 49:362–365.
43. Rajashekar B, et al. (2009) Expansion of signal pathways in the ectomycorrhizal fungus *Laccaria bicolor*—evolution of nucleotide sequences and expression patterns in families of protein kinases and RAS small GTPases. *New Phytol* 183:365–379.
44. Doddapaneni H, Chakraborty R, Yadav JS (2005) Genome-wide structural and evolutionary analysis of the P450 monooxygenase genes (P450ome) in the white rot fungus *Phanerochaete chrysosporium*: Evidence for gene duplications and extensive gene clustering. *BMC Genomics* 6:92.
45. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
46. Kämper J, et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97–101.
47. Loftus BJ, et al. (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307:1321–1324.
48. Hosack DA, Dennis G, Jr, Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol* 4:R70.
49. Saeed AI, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
50. Fisher R (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford, UK).
51. Kimura M (1956) A model of a genetic system which leads to closer linkage by natural selection. *Evolution* 10:278–287.
52. Nei M (1967) Modification of linkage intensity by natural selection. *Genetics* 57: 625–641.
53. Pál C, Hurst LD (2003) Evidence for co-evolution of gene order and recombination rate. *Nat Genet* 33:392–395.
54. Wong S, Wolfe KH (2005) Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet* 37:777–782.
55. Swamy S, Uno I, Ishikawa T (1984) Morphogenic effects of mutations at the *A* and *B* incompatibility factors in *Coprinus cinereus*. *J Gen Microbiol* 130:3219–3224.
56. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98:5116–5121.
57. Sanderson MJ (2003) r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
58. James TY, et al. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443:818–822.
59. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: Review and new analyses. *Mycologia* 98:838–849.
60. Calabrese PP, Chakravarty S, Vision TJ (2003) Fast identification and statistical evaluation of segmental homologies in comparative maps. *Bioinformatics* 19(Suppl 1):i74–i80.

Fig. S1. Summary plots of the 13 chromosomes of *C. cinerea*. Each plot shows the location of (Top panel) telomeres (if identified) in red, and centromere as a black oval; (Second panel) the density of transposable elements (brown); (Third panel) tRNA genes (light green); (Fourth panel) recombination rates (the position of the SSR markers is indicated by vertical black bars, and white is unmapped; red is high recombination; gray is average recombination; blue is low recombination); (Fifth panel) the density of all genes (orange); (Sixth panel) the density of orphan genes (light orange); (Seventh panel) the density of orthologous genes (blue); (Eighth panel) the density of paralogous genes (red); (Ninth panel) similarity of paralog families represented as 1/dS (synonymous substitution rate); (10th panel) syntenic regions (all regions of synteny between *C. cinerea* and *L. bicolor* are indicated in green, and blocks with >15 anchors are indicated in dark green). Vertical scales are defined for each bar in the bar title. Horizontal scale is Mb (chromosome XIII is 0.982 Mb).

Fig. S1.

Fig. S2. Chromosomal distribution of protein kinases. Regions of high recombination (red), low recombination (blue), and extensive synteny (light gray) as in Fig. S1. Green bars indicate protein kinases from widely conserved groups. Orange bars indicate FunK1 protein kinases.

Fig. S2.

Fig. S3. Phylogenetic analysis of the MAPK genes across nine fungi. Arrows indicate Basidiomycete-specific duplications. The prefixes of genes in the tree indicate species: AFUA, *Aspergillus fumigatus*; BDEG, *Batrachochytrium dendrobatidis*; CC, *C. cinerea*; CNAG, *C. neoformans* serotype A; LACBI, *L. bicolor*; SCCOM, *Schizophyllum commune*; SP, *Schizosaccharomyces pombe*; UM, *U. maydis*; Y, *S. cerevisiae*. Numbers at the internal nodes represent the posterior probability for the clade from MrBayes.

Fig. S3.

Fig. S4. Phylogenetic analysis of TRIBE-MCL cluster 75 containing a subfamily of the P450 gene family. The tree shows that both local duplication (stars and florets indicate sets of adjacent gene pairs) and independent lineage-specific duplications have shaped the evolution of this family. Numbers at the internal nodes represent the posterior probability for the clade from MrBayes with thickened branches scaled to indicate more significant support.

Fig. S4.

Fig. S5. Phylogenetic analysis of the hydrophobin gene family. The MrBayes-computed tree depicts the hydrophobin domain containing genes in *C. cinerea* (prefixed by CC1G), *L. bicolor* (prefixed by Lbic), and *P. chrysosporium* (prefixed by Pchr) rooted with the *U. maydis* hydrophobin gene UM05010. Stars, crosses, and florets indicate clusters of adjacent genes (separated at most by one gene), with the two sets found on chromosome X indicated by a labeled vertical bar. Numbers at the internal nodes represent the posterior probability for the clade from MrBayes with thickened branches scaled to indicate more significant support.

Fig. S5.

Fig. S6. Structure of the A and B mating type loci. (*A*) The A locus. (*B*) The B locus. In both A and B, conserved sequence regions are indicated by a bold line; unique sequences of the Aα, Aβ, and B subloci are indicated by a thin line. Exons are shown in light blue. The 5′ UTR regions are shown in green, and the 3′ UTR regions are shown in red. Numbers indicate gene start and stop positions. Direction of transcription and gene names are indicated below the line. Paralogous duplications are shown in the same color.

Fig. S6.

Other Supporting Information Files

Dataset S1 (XLS)