

## Evolution of the Mating Type Locus: Insights Gained from the Dimorphic Primary Fungal Pathogens *Histoplasma capsulatum*, *Coccidioides immitis*, and *Coccidioides posadasii*<sup>∇†</sup>

James A. Fraser,<sup>1</sup> Jason E. Stajich,<sup>2</sup> Eric J. Tarcha,<sup>3</sup> Garry T. Cole,<sup>3</sup> Diane O. Inglis,<sup>4</sup>  
Anita Sil,<sup>4</sup> and Joseph Heitman<sup>5\*</sup>

School of Molecular and Microbial Sciences, University of Queensland, Brisbane, QLD 4072, Australia<sup>1</sup>; Department of Plant and Microbial Biology, University of California, Berkeley, California 94720<sup>2</sup>; Department of Biology, University of Texas at San Antonio, San Antonio, Texas 78249<sup>3</sup>; Department of Microbiology and Immunology, University of California, San Francisco, California 94143<sup>4</sup>; and Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710<sup>5</sup>

Received 17 January 2007/Accepted 14 February 2007

**Sexual reproduction of fungi is governed by the mating type (*MAT*) locus, a specialized region of the genome encoding key transcriptional regulators that direct regulatory networks to specify cell identity and fate. Knowledge of *MAT* locus structure and evolution has been considerably advanced in recent years as a result of genomic analyses that enable the definition of *MAT* locus sequences in many species as well as provide an understanding of the evolutionary plasticity of this unique region of the genome. Here, we extend this analysis to define the mating type locus of three dimorphic primary human fungal pathogens, *Histoplasma capsulatum*, *Coccidioides immitis*, and *Coccidioides posadasii*, using genomic analysis, direct sequencing, and bioinformatics. These studies provide evidence that all three species possess heterothallic bipolar mating type systems, with isolates encoding either a high-mobility-group (HMG) domain or an  $\alpha$ -box transcriptional regulator. These genes are intact in all loci examined and have not been subject to loss or decay, providing evidence that the loss of fertility upon passage in *H. capsulatum* is not attributable to mutations at the *MAT* locus. These findings also suggest that an extant sexual cycle remains to be defined in both *Coccidioides* species, in accord with population genetic evidence. Based on these *MAT* sequences, a facile PCR test was developed that allows the mating type to be rapidly ascertained. Finally, these studies highlight the evolutionary forces shaping the *MAT* locus, revealing examples in which flanking genes have been inverted or subsumed and incorporated into an expanding *MAT* locus, allowing us to propose an expanded model for the evolution of the *MAT* locus in the phylum *Ascomycota*.**

In the fungal kingdom, sexual reproduction is regulated by a specialized genomic region known as the mating type (*MAT*) locus (12, 13). This important genomic feature has been most extensively studied in the largest of the fungal phyla, the *Ascomycota*. Most commonly, the sexual members of this phylum have a bipolar mating type system, where strains are one of two mating types. For mating to occur in these heterothallic species, cells of differing mating types must come together. In those species where the *MAT* locus has been characterized at a molecular level, this process is regulated by a bipolar system: cells normally carry one of two different alleles of the *MAT* locus. These alleles, known as idiomorphs, contain unrelated sequences that encode different transcription factors. In the euascomycetes, which include the medically important dimorphic pathogens and the majority of the ascomycete molds, one *MAT* locus allele encodes a high-mobility-group (HMG) domain-type transcription factor, and the other allele encodes an  $\alpha$ -box domain transcription factor. When

two isolates of a species with differing *MAT* loci come together under appropriate conditions, sexual reproduction can proceed. Importantly, for many ascomycete species, a sexual cycle has never been observed, yet analysis of the genomic sequence has revealed the presence of potentially functional mating type loci.

Over the past decade, our understanding of the structure of fungal mating type loci has blossomed, and elegant studies have begun to elucidate the evolutionary processes that fashioned these unusual genomic structures (4, 11, 18, 35). Despite the quantity of information available regarding *MAT* locus structures, there is a paucity of information regarding the *Onygenales*, one of the most medically important orders of fungi. The members of the *Onygenales* include a number of dimorphic primary pathogens, including *Histoplasma capsulatum*, *Coccidioides immitis*, and *Coccidioides posadasii*, all capable of causing life-threatening systemic mycoses (6, 38). Although the *MAT* locus has not been characterized in these species, there is strong evidence that functional loci exist and that they play important roles in the life cycle of these organisms.

While sexuality is uncommon among pathogenic fungi, *H. capsulatum* is one of the exceptions to this generalization. This primary pathogen is distributed worldwide in association with soil enriched with bird and bat guano and is thought to represent multiple cryptic species (19). In the soil, the mycelial

\* Corresponding author. Mailing address: Department of Molecular Genetics and Microbiology, 322 CARL Building, Research Drive, Duke University Medical Center, Durham, NC 27710. Phone: (919) 684-2824. Fax: (919) 684-5458. E-mail: heitm001@duke.edu.

† Supplemental material for this article may be found at <http://ec.asm.org/>.

<sup>∇</sup> Published ahead of print on 2 March 2007.

phase asexually produces microconidia that are inhaled into the lung, causing histoplasmosis (14, 34, 37). The heterothallic bipolar mating system of *H. capsulatum* was first described in 1972, with strains designated either + or - (23, 24), and represents one of the first discoveries of the sexual structures of a human fungal pathogen. Fertility is lost rapidly during laboratory passage, implying that selective pressures may serve to maintain fecundity in the environment (26).

Despite the prevalence of both mating types of this fungus (environmental samples exhibit a 1:1 ratio of the two mating types), in clinical specimens highly skewed ratios of 7:1 (-/+) have been observed (25, 26). Remarkably, murine infection experiments revealed equivalent virulence potentials between isolates of opposite mating types, suggesting that + and - isolates may therefore differ in the ability to produce the infectious propagule. The - strains have been found to display an increased ability to convert from hyphae to yeast, the cell type that is more commonly the cause of disease in humans (22, 26). How mating type may regulate this dimorphic transition and increase virulence is unknown.

In contrast to *H. capsulatum*, which is distributed worldwide, the *Coccidioides* species are restricted to the Americas (32). The soilborne saprobic multicellular hyphal form of this fungus septates into single-celled arthroconidia that can become airborne (such as during dust storms) and be inhaled to produce life-threatening infections in otherwise healthy individuals (6). Although a sexual cycle has never been observed in the laboratory, molecular phylogenetic analysis revealed evidence supporting sexual reproduction (3, 21). These studies were key in revealing two extremely closely related yet independent species, *C. immitis* and *C. posadasii*, which are distinct from each other and represent discrete populations undergoing independent sexual recombination (9, 20, 21). While these studies provide evidence that *C. posadasii* and *C. immitis* undergo a sexual cycle in nature, this has never been observed in the laboratory. Furthermore, as neither mating nor the *MAT* locus has been identified, it is unknown whether different mating types have distinct pathogenic capacities, such as has been seen for *H. capsulatum*.

The structure of the *MAT* loci of the fungi of the order *Onygenales* is therefore an important topic to address, both relative to the role of these fungi as pathogens and from an evolutionary perspective. Here, we describe an analysis of the *MAT* locus structure of the dimorphic pathogens *H. capsulatum*, *C. posadasii*, and *C. immitis* by combining bioinformatic analyses of genomes, direct sequence confirmation, PCR analysis of multiple isolates of each species, and comparative analyses with other fungi of the phylum *Ascomycota*. Our findings have revealed that all three species have canonical euascomycete bipolar *MAT* locus structures, with strains known or predicted to be of opposite mating types encoding either an HMG domain or an  $\alpha$ -box domain transcription factor at this locus. Potential mating partners have been identified for both *Coccidioides* species, and the + and - mating types of *Histoplasma* have been identified as belonging to the  $\alpha$ -box and HMG types, respectively. Finally, this work provides critical information contributing to an extended model of *MAT* locus evolution in the most populated of the fungal phyla.

## MATERIALS AND METHODS

**Strains and media.** The reference strains used in this study were *H. capsulatum* strains G217B, ATCC 22635, ATCC 22636, G186AR, and G217B *ura5-23*, *C. immitis* strains RS and H538.4, and *C. posadasii* strains Silveira, 1037, C735, and C634. *C. immitis* strain H538.4 is an environmental isolate, while the other isolates of *Coccidioides* are clinical derivatives. Strains were grown under modified biosafety level 2 (BSL2) or BSL3 conditions as appropriate. *H. capsulatum* strains ATCC 22635 (*MAT1-2/-*) and ATCC 22636 (*MAT1-1/+*) were received directly from the American Type Culture Collection as mycelial frozen stocks and grown on soil extract agar at room temperature and on *Histoplasma* macrophage medium plates at 37°C under 5% CO<sub>2</sub> (39).

**Genomic DNA isolation.** For *H. capsulatum* genomic DNA isolation, 100 ml of yeast culture of strains ATCC 22635 and ATCC 22636 was grown in *Histoplasma* macrophage medium at 37°C under 5% CO<sub>2</sub> at 150 rpm for 3 days. Genomic DNA was isolated using a QIAGEN genomic DNA kit according to the manufacturer's instructions. The genomic DNA was isolated from *Coccidioides immitis* and *Coccidioides posadasii* mycelia grown for 4 days in glucose-yeast extract liquid medium (Difco) as previously described (30).

**Sequencing, assembly, and bioinformatics.** Sequencing reactions were performed with an MJ Research thermal cycler using standard BigDye Terminator chemistry (Applied Biosystems) and analyzed on a PE3700 96-capillary sequencer. Sequence reads were assembled using Sequencher (Gene Codes Corporation, Ann Arbor, MI). Additional analysis of the data was performed using BLAST (1) and MacVector (MacVector Inc., Cary, NC). Based on the initial assembly of the end sequences, oligonucleotides were selected to close gaps in the sequence coverage by primer walking. Genes were annotated based on homology to the existing annotation in GenBank. Dot plots were generated using DOTUP from the EMBOSS package.

**PCR.** Primers JOHE13618 and JOHE13619 were used to amplify the entire *H. capsulatum* *MAT* locus. Primers JOHE13950 and JOHE13951 were used to amplify the *C. immitis* and *C. posadasii* *MAT1-2* idiomorph. Primers JOHE18241 to JOHE18248 were used for *MAT* diagnostics. PCR analysis was performed using the ExTaq polymerase blend (Miris) with primers given in the following list: JOHE13618, 5'-AGG CCA TAA CAC TGA CCG TAT-3'; JOHE13619, 5'-GCA TTA CAT CGC AGC ACA TTG-3'; JOHE13950, 5'-GTC ACG AAG AAA CCC GGA ATC-3'; JOHE13951, 5'-AAC AAA CTT TCG CAG GCA AGG-3'; JOHE18241, 5'-TCG TCT ACT TTG GAC TTC GGA C-3'; JOHE18242, 5'-CTG CTA TTG CTT GCT CTG AAC C-3'; JOHE18243, 5'-CAG CCA ATG ACT GGT TCT AAG G-3'; JOHE18244, 5'-TTT ACG GGG AGC ATA CTG GTA G-3'; JOHE18245, 5'-AGG AAA CGA TGT CTC TGC CGT C-3'; JOHE18246, 5'-TAC ACG AAG GTA ATC ACT TGG G-3'; JOHE18247, 5'-AAC AAG CAA TGA CCA AAG CGT C-3'; and JOHE18248, 5'-AGG TGA GAG CGG AAC AAT GAG G-3'.

***H. capsulatum* mating assays.** ATCC 22635 and ATCC 22636 were prepared for mating tests by inoculating mycelia from soil agar plates into Sabouraud dextrose broth (Difco). The mating of each of these strains was also tested against that of two common laboratory strains, G217B (*MAT1-1/+*) and G186AR (*MAT1-2/-*). Two different G217B isolates were used in mating tests. The first isolate was originally a kind gift of William Goldman to the Sil laboratory and has undergone significant passaging. The second G217B isolate was obtained directly from the American Type Culture Collection (ATCC 26032) and presumably has not been passaged as extensively. Mating between ATCC 22635 and ATCC 22636 and the G217B and G186AR strains was tested by streaking pairwise combinations of each strain on AlphaCel agar (MP Biochemicals), sealing the plates with parafilm, and incubating the plates at 25°C. Self-mating controls for each strain were also performed. After 3 weeks, the plates were observed by eye and microscopically for the presence of perithecia or any type of mating structure. However, no mating structures were observed, even after several months of incubation.

**Diagnostic amplification of the *Histoplasma* and *Coccidioides* predicted *MAT* regions.** PCR amplification was conducted on each of six *Coccidioides* DNA templates, five *H. capsulatum* templates, and one negative control for each pair of primers. For *H. capsulatum*, *MAT1-1* ( $\alpha$ -box)-specific PCR employed primers JOHE18245 and JOHE18246 while *MAT1-2* (HMG domain)-specific PCR employed primers JOHE18247 and JOHE18248. For *Coccidioides*, *MAT1-1* ( $\alpha$ -box)-specific PCR employed primers JOHE18241 and JOHE18242 while *MAT1-2* (HMG domain)-specific PCR employed primers JOHE18243 and JOHE18244. Thermal cycling consisted of a 2-minute denaturation step at 95°C, followed by 30 cycles of denaturation at 94°C for 15 seconds, annealing at 55°C for 15 seconds, and extension at 72°C for 3 min and 30 seconds. After the completion of the 30 cycles, the samples were held at 72°C for 10 min. Amplicons were resolved by gel electrophoresis.

**Phylogenetic analyses.** For the species tree (see Fig. 5), the best reciprocal orthologs were identified via InParanoid (31) and combined with single-linkage clustering to identify gene sets with a single member per species producing 662 orthologous gene families. Protein alignments were constructed with ProbCons (8) and filtered with Gblocks (5), and phylogenetic trees were inferred with MrBayes (33) for each gene family. Gene trees were built with the topological constraint of *Schizosaccharomyces pombe* to be basal in the ascomycetes and with the hemiascomycetes and euascomycetes as a monophyletic group. A consensus supertree representing the putative species tree was built from the individual gene trees by using Clann (7). A species tree was also built with MrBayes by using a concatenated alignment of codons of 100 randomly selected ortholog families. The consensus supertree and the concatenated data set provided identical topologies.

*MAT* locus gene trees were constructed from ProbCons alignments of the predicted proteins and filtered for best-aligned blocks with Gblocks, and phylogenetic trees were inferred and bootstrapped with MrBayes and PHYML (5, 8, 16, 33). Trees were visualized with TreeViewX (29) and edited with Adobe Illustrator.

**Nucleotide sequence accession numbers.** The novel mating type locus sequences have been submitted to GenBank under accession numbers EF472255 and EF472256 (*H. capsulatum*) and EF472257, EF472258, and EF472259 (*C. immitis* and *C. posadasii*).

## RESULTS AND DISCUSSION

**Candidate mating type locus in *Histoplasma*.** Despite the paucity of information regarding the *Onygenales* mating type loci, significant advances have recently been made regarding the most closely related order, *Eurotiales*. With the aid of genomic studies, the *MAT* loci of three *Aspergillus* species have been defined. In the opportunistic pathogen *Aspergillus fumigatus* and the industrially used *Aspergillus oryzae*, the *MAT* loci encode an  $\alpha$ -box domain transcription factor or an HMG domain transcription factor (15). These are not simply alleles of a gene that have diverged over evolutionary time; rather, they are unrelated sequences that have been termed “idiomorphs” (27). Using the nomenclature defined by Turgeon and Yoder (36), due to the types of transcription factors they encode, these alleles are defined as the *MAT1-1* allele and the *MAT1-2* allele, respectively. In both cases, these loci are flanked by the *APN2* (encoding DNA lyase) and *SLA2* (encoding a cytoskeletal protein) genes. In the homothallic species *A. nidulans*, both *MAT* alleles are present; however, the flanking regions have been altered due to a translocation. In this case, the *A. nidulans*  $\alpha$ -box locus is associated with *SLA2*, while the HMG domain locus is associated with *APN2*.

BLAST analysis of the incomplete *H. capsulatum* Wu24 genome (Broad Institute) by using *A. fumigatus* *SLA2* and *APN2* reveals that they lie adjacent to a predicted *MAT1-1* allele encoding an  $\alpha$ -box transcription factor homolog. The structure of this *MAT* locus is consistent with a classic ascomycete bipolar system in which the presumed second mating type encodes a *MAT1-2* HMG domain idiomorph. Although the content of the predicted *MAT* alleles was equivalent to that seen in aspergilli, there are minor changes in the structure of this genomic region in *H. capsulatum* (Fig. 1). In aspergilli, the *APN2* and *SLA2* genes are convergently transcribed toward the *MAT* locus. In contrast, in *H. capsulatum*, the *APN2* and *MAT*-distal *COX13* genes (encoding the cytochrome *c* oxidase subunit VIa homolog) have been inverted, changing the local gene order from *COX13-APN2-MAT-SLA2* to *APN2-COX13-MAT-SLA2*.

Importantly, although mating has been observed in *H. capsulatum*, the mating type of strain Wu24 was unknown. To

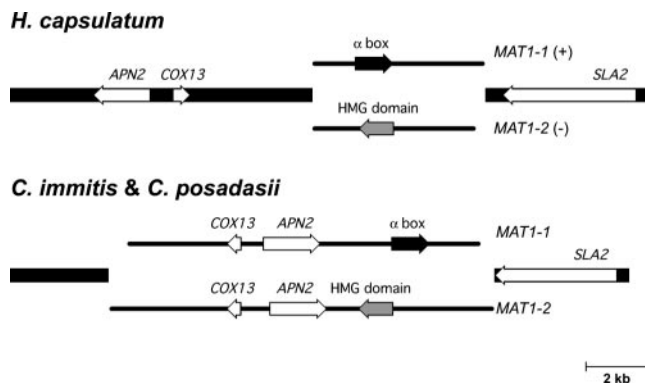


FIG. 1. Structure of the *MAT* locus in primary pathogens from the *Onygenales*. Sequencing and bioinformatic analyses of genomic sequences spanning the mating type loci of *H. capsulatum*, *C. immitis*, and *C. posadasii* have revealed them to encode *MAT* idiomorphs very similar in structure to those seen in aspergilli. Left and right flanking regions are depicted as thick lines; alternate *MAT*-specific sequences are depicted as thin lines.

determine the mating type relative to the original studies that identified the +/- system, we performed a PCR analysis of the original *H. capsulatum* mating strains ATCC 22635 (-) and ATCC 22636 (+) by using primers anchored in the *APN2* and *SLA2* genes, respectively, flanking the predicted locus (JOHE13618 and JOHE13619). Using this approach, fragments of different lengths were obtained and sequenced using primer walking. Analysis of the amplified *MAT* regions revealed that the + *MAT* idiomorph of ATCC 22636 is 5,418 bp in length and encodes an  $\alpha$ -box protein. In contrast, the - *MAT* idiomorph of ATCC 22635 is 5,088 bp in length and encodes a predicted HMG domain protein (Fig. 1). Inspection of the open reading frames encoded by the  $\alpha$ -box and HMG domain *MAT* genes revealed that both genes are apparently intact, with no evidence of loss-of-function mutations in any of the strains analyzed, consistent with previous findings that this species exhibits an extant heterothallic bipolar sexual cycle.

While we attempted to verify the linkage of these loci with mating type, we were unable to recapitulate mating in the isolates of this pathogen that were available to us for testing (see Materials and Methods). In accordance with the unified nomenclature system of the euascomycetes, *H. capsulatum* + isolates are therefore *MAT1-1*, and - strains are *MAT1-2*. Accordingly, we designate the Wu24 genome strain as being mating type +. Furthermore, these findings confirm that the *APN2-COX13-MAT-SLA2* gene order is conserved in both mating types. Subsequent sequencing of additional *H. capsulatum* isolates at Washington University in St. Louis revealed similar findings, with BLAST analyses showing strain G186AR to be mating type - and strain G217B to be mating type +.

Due to the importance of the mating type on rates of infection, we employed the identified sequences to design diagnostic primer sets to enable the rapid identification of the mating type in this primary pathogen. Primers were designed to amplify either the *MAT1-1*- or the *MAT1-2*-encoded transcription factor. Using a basic thermocycling program, we were reproducibly able to amplify the *MAT*-encoded transcription factor genes, effectively enabling the rapid determination of mating type (Fig. 2).

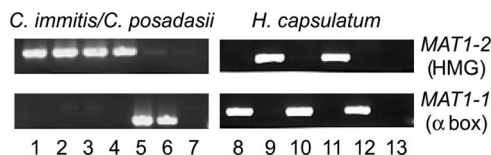


FIG. 2. Results of a diagnostic PCR of the *MAT* locus of *Histoplasma* and *Coccidioides*, establishing molecular mating type allele specificity by PCR for *C. immitis*, *C. posadasii*, and *H. capsulatum*. Genomic DNA was subject to PCR amplification with species-specific primer pairs for the HMG or  $\alpha$ -box genes in the *MAT* locus. The strains analyzed in the left panels were *C. immitis* RS (lane 1) and H538.4 (lane 5) and *C. posadasii* Silveira (lane 2), 1037 (lane 3), C735 (lane 4), and C634 (lane 6). The strains analyzed in the right panels were *H. capsulatum* G217B (lane 8), ATCC 22635 (lane 9), ATCC 22636 (lane 10), G186AR (lane 11), and G217B *ura5* (lane 12). "No DNA" controls were included in lanes 7 and 13. The upper panels represent reactions with primers specific for the HMG domain gene, and the lower panels represent reactions with primers specific for the  $\alpha$ -box gene. Products correspond to 549 bp (*Coccidioides*  $\alpha$  box), 689 bp (*Coccidioides* HMG domain), 592 bp (*H. capsulatum*  $\alpha$  box), and 564 bp (*H. capsulatum* HMG domain).

**Candidate mating type locus in *Coccidioides*.** Extending our analysis to the related primary pathogens *C. immitis* and *C. posadasii*, we searched the genomes in progress at the Broad Institute and The Institute for Genomic Research (genome strains RS and C735, respectively). In contrast to that observed with our original search of the Wu24 *H. capsulatum* genome, BLAST analysis using *A. fumigatus* *SLA2* and *APN2* revealed that they lie adjacent to a predicted HMG domain-type *MAT1-2* allele (Fig. 1). Comparison of the two sibling species reveals that the encoded proteins are almost identical, with only three changes across the predicted 339-amino-acid protein. Again, the structure of this *MAT* locus is consistent with a classic ascomycete bipolar system, except in this case, the presumed second mating type encodes a *MAT1-1*  $\alpha$ -box idiomorph. Once again, primers anchored in the *APN2* and *SLA2* genes flanking the predicted locus (JOHE13950 and JOHE13951, respectively) were employed to amplify this genomic region from alternative strains in an effort to identify the opposite mating type.

Using this approach, ~5-kb fragments were obtained from the control *C. immitis* strain RS sample in addition to *C. posadasii* strains Silveira and 1037. Sequencing of the amplified fragments from *C. posadasii* strains Silveira (GenBank accession no. EF472258) and 1037 (GenBank accession no. EF472257) showed that each encodes almost identical *MAT1-2* HMG domain alleles, as in the *C. posadasii* genome strain C735. However, no product was obtained for *C. immitis* strain H538.4 or *C. posadasii* strain C634, suggesting that they may correspond to *MAT1-1*  $\alpha$ -box isolates. Despite repeated attempts, these strains remained recalcitrant and no PCR product could be detected.

During the course of this work, additional *C. immitis* genomes for strains H538.4 and RMSCC2394 were sequenced at the Broad Institute. Subsequent bioinformatic analyses of these isolates revealed that like the previous two *Coccidioides* genomes, RMSCC2394 is a *MAT1-2* HMG domain isolate. In contrast, H538.4, one of the two strains from which we were unable to amplify *MAT* by using a PCR-based approach, encoded a *MAT1-1*  $\alpha$ -box idiomorph as we had hypothesized. Employing this new sequence, we designed *Coccidioides*-spe-

cific diagnostic primer sets to amplify either the *MAT1-1*- or the *MAT1-2*-encoded transcription factor and facilitate mating type determination (Fig. 2). In addition to confirming the known mating types of *C. immitis* strains RS and H538.4 and *C. posadasii* strains Silveira and 1037, this approach revealed that the previously untypeable *C. posadasii* strain C634 corresponded to the missing *MAT1-1* mating type.

Inspection of the open reading frames encoded by the  $\alpha$ -box and HMG domain *MAT* genes revealed that both genes are apparently intact, with no evidence of loss-of-function mutations in any of the strains analyzed. This is consistent with previous population genetic studies indicative of sexual recombination and provides additional support for the hypothesis that an extant heterothallic bipolar sexual cycle remains to be defined for both species. The identification of strains of opposite mating types in both *C. immitis* and *C. posadasii* now provides a platform from which to initiate attempts to characterize the sexual cycle in these primary pathogens, an aspect of the life cycle that is relevant both as a tool in the laboratory and from an epidemiological perspective.

**A step in the evolution of sex chromosomes: expansion of a *MAT* locus.** Following the identification of *Coccidioides* *MAT1-1* isolates, we sought an explanation for why the original PCR-based efforts were unable to identify this second mating type. Comparison of the genomic regions surrounding the *C. immitis* and *C. posadasii* *MAT* loci showed 100% synteny with the gene order in *A. fumigatus* and *A. oryzae* (Fig. 1). This also holds true when we compare the structures found in the genome project at the Broad Institute of the next closest sequenced relative, *Uncinocarpus reesii* (30), which encodes a *MAT1-1*  $\alpha$ -box idiomorph. As seen for *C. immitis*, *C. posadasii*, *A. fumigatus*, and *A. oryzae*, this member of the *Onygenales* has the *COX13-APN2-MAT-SLA2* mating type locus region arrangement. It was not until a pairwise comparison was performed between the *MAT1-1* and *MAT1-2* idiomorphs of *Coccidioides* that the most significant changes to this locus became apparent. In contrast to the other species, in *Coccidioides*, the dissimilarity between the two idiomorphs expands beyond the HMG domain and  $\alpha$ -box regions to encompass the adjacent *APN2* and *COX13* gene pair (Fig. 3). Analysis of the amplified *MAT* regions revealed them to be atypically large for this type of structure in the *Ascomycota*, with the expanded *MAT1-1* idiomorph of strain H538.4 being 8,062 bp in length while the *MAT1-2* idiomorph of strain RS is 9,037 bp (Fig. 1). This is a clear example of the acquisition of genes adjacent to *MAT* being incorporated into the locus, expanding this important region by increasing both size and gene content. There are direct parallels between what we see here and the model we have previously proposed for the expansion of the *MAT* locus of the basidiomycete pathogen *Cryptococcus neoformans*, whose *MAT* locus, which is >100 kb in length and encodes more than 25 genes, is even larger than that of *Coccidioides* (11).

Remarkably, in this case, there is little evidence to suggest that the formation of *MAT*-specific alleles of captured genes is a slow and gradual process. Phylogenetic analysis of the *APN2* gene showed that contrary to this gene forming closely related yet distinct *MAT1-1* and *MAT1-2* clades, these clades were instead dramatically different. When the *C. immitis* *MAT1-2* alleles are not considered, this gene shows phylogeny as is

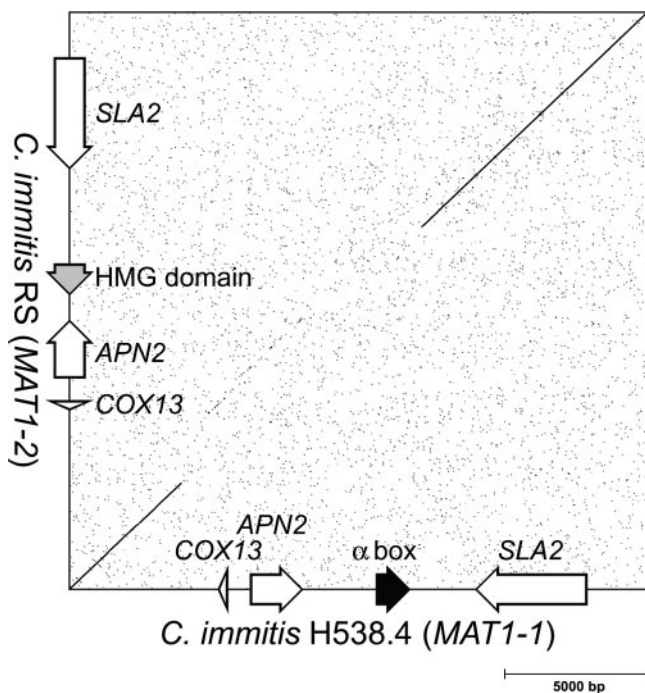


FIG. 3. Expansion of a *MAT* locus. Pairwise comparison of the *MAT1-1* and *MAT1-2* alleles of *Coccidioides* by using DOTTUP reveals that the mating type locus has expanded, entrapping the *COX13* and *APN2* genes that normally occupy the left-hand boundary of the locus. This has created a hybrid locus, consisting of both idiomorphic segments (encoding the  $\alpha$ -box and HMG domain transcription factors) and divergent alleles of these two captured genes.

expected for any gene in these species (Fig. 4). In contrast, the *C. immitis* and *C. posadasii* *MAT1-2* *APN2* alleles form a clade basal and separate from that of the rest of the euascomycetes. How such a relationship has developed is unknown, with this example almost resembling a case of horizontal gene transfer from a distantly related species. Analysis with PAML comparing the *APN2* *MAT1-1* and *MAT1-2* gene sequences indicates that the sequences have been under positive selection, with the targets of selection being sites in the region between the endonuclease and the zinc-finger domain (see Fig. S1 in the supplemental material). These results warrant further analysis of the functional consequences of the changes in the *APN2* genes to understand what effects the changes have on the activity of the protein. As none of the significant changes appear to occur in either of the conserved functional domains, perhaps the changes have a role in a new conformational structure. In contrast to the *APN2* gene, the *COX13* gene, which has also been incorporated into *MAT*, has been subject to more-gradual divergence of the two *MAT* alleles or more-recent gene conversion (data not shown).

The *Coccidioides* *MAT* locus therefore represents a transitional form of the mating type locus and has evolved beyond the classical ascomycete idiomorphic form to exist instead as a type of hybrid structure. In addition to the continued presence of the idiomorphic region, since its divergence from *Histoplasma*, this structure now has a clearly delineated section that contains divergent alleles of two genes. This discovery helps to explain aspects of the unusual locus of *Candida albi-*

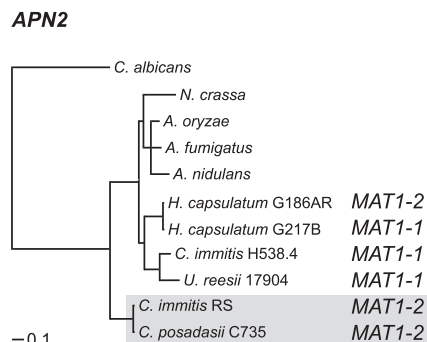


FIG. 4. Phylogeny of the *APN2* gene. In *C. immitis* and *C. posadasii*, the *APN2* gene has been captured by the *MAT* locus. Contrary to a model of gradual divergence into *MAT*-specific alleles, phylogenetic analysis shows that the *APN2* allele present in the *MAT1-2* locus is dramatically different, resembling instead a gene from outside the euascomycetes. The scale bar represents 0.1 substitutions per site.

*cans*, where, in addition to the mating type transcription factors, three additional genes are now present (17, 35). These may have been enveloped by *MAT* in the same way as has occurred in *Coccidioides*. In the case of *C. albicans*, the genes encode a phosphatidylinositol 4-kinase, a poly(A) polymerase, and an oxysterol binding protein, which are unrelated to the genes that have been incorporated into the mating type locus of *Coccidioides*. Interestingly, these genes have not been incorporated into the *MAT* locus of *H. capsulatum* but instead have been retained as flanking sequences. Other examples of novel genes resident in the *MAT* locus are also known. These include *Neurospora crassa* and, most poignantly, the basidiomycetes *C. neoformans* and *Cryptococcus gattii*, where the *MAT* locus has expanded to span more than 100 kb and more than 25 genes (at least some of which do not function directly in sexual reproduction). Whether the incorporation of originally flanking genes into the *Coccidioides* *MAT* locus, and their subsequent divergence, contributes to cell identity or sexual reproduction in *Coccidioides* or is simply a consequence of their close juxtaposition to the ancestral *MAT* locus remains to be tested experimentally.

**Evolution of the *MAT* locus in ascomycetes.** The findings presented here provide a molecular description of the mating type locus for three thermally dimorphic fungal species, *Histoplasma capsulatum*, *Coccidioides immitis*, and *Coccidioides posadasii*, all of which are primary pathogens of humans. A sexual cycle has been previously described only for *H. capsulatum* (23, 24), but population genetic studies suggest that cryptic sexual reproduction is likely to occur in *Coccidioides* as well (21). The fact that the two alleles in *H. capsulatum* encode either an HMG domain protein or an  $\alpha$ -box protein, and that these alleles are present only once in the genome, is in full accord with the description of *H. capsulatum* as having a heterothallic sexual cycle. That the *MAT* loci of *C. immitis* and *C. posadasii* are arranged similarly allows the predication that the sexual cycle will also be heterothallic in these two pathogenic species. Importantly, none of the three species contain fused *MAT* loci or evidence of both alleles present within the genome, and thus, all three are molecularly distinguished from homothallic species that have been defined in the fungal king-

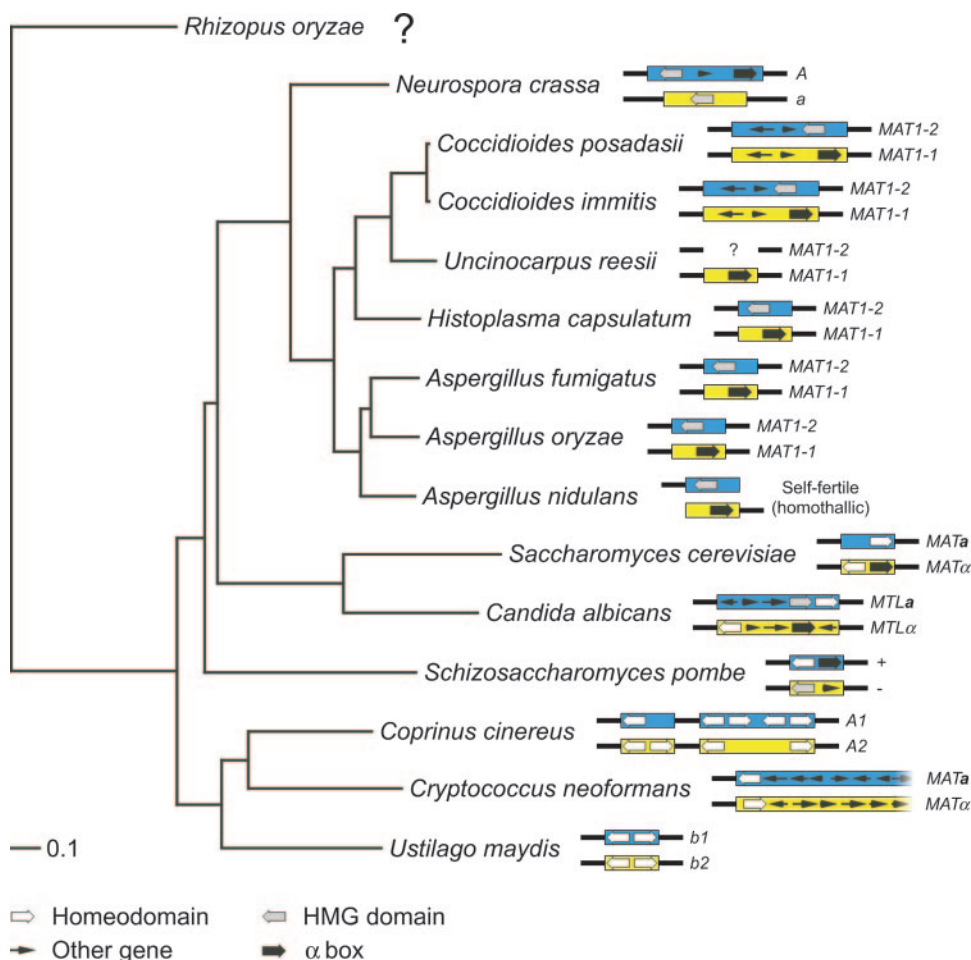


FIG. 5. Evolution of *MAT* in euscomycetes. Analysis of *MAT* locus structure relative to the phylogenetic relationship has revealed the evolutionary steps that have shaped the formation of these genomic sex-determining regions in fungi. Building upon previous studies by Tsong et al. (35) and Butler et al. (4), the addition of sequences from the *Onygenales* (including *H. capsulatum*, *U. reesii*, *C. immitis*, and *C. posadasii*) reveals that the common ancestor of the *Onygenales* and aspergilli likely had equivalent *MAT* idiomorphs also flanked by *APN2* and *SLA2*. Mating type locus structures are not to scale. The consensus supertree was constructed as described by Fitzpatrick et al. (10).

dom, including those with silent mating type cassettes in which mating type switching can occur. Comparison of the *MAT* loci of *C. immitis* and *C. posadasii* with that of *Uncinocarpus reesii*, the closest extant species with a defined sexual cycle, further supports the assignment of *MAT* in the pathogenic sibling species and buttresses the hypothesis that an extant sexual cycle remains to be defined.

Our analysis of the gene structure of all the sequenced *MAT* alleles provides no evidence of loss-of-function mutations in either the HMG domain or the  $\alpha$ -box gene, indicating that these may still contain functional alleles and have not been subject to the pseudogene formation that would be predicted to occur if these species had evolved to be asexual. The molecular definition of the *MAT* locus and the description of molecular reagents that can be used to establish the mating types of strains by rapid PCR analysis provide approaches that can now be applied to understand some of the mysteries surrounding these pathogens. These include the interesting finding that sexual fecundity is rapidly lost during laboratory passage of *H. capsulatum* isolates, the link between mating type

and virulence potential in *H. capsulatum*, and the role that mating or mating type may have played in the speciation and geographic expansion of *Coccidioides* species from North America to Central America and South America. Perhaps most significantly, these discoveries can also be applied to the generation of congenic strain pairs for genetic and virulence studies of these dangerous pathogens.

The elucidation of the molecular nature of the *MAT* locus for *H. capsulatum*, *C. immitis*, and *C. posadasii* also provides considerable insight into the evolution of mating type loci in fungi (Fig. 5). These studies reveal that the two idiomorphic *MAT* alleles in all three species encode either an HMG domain or an  $\alpha$ -box transcription factor, similar to those of other euscomycete fungal species in which the *MAT* locus has been characterized. A central question is the molecular nature of the ancestral *MAT* locus in the fungal kingdom.

Ohno first hypothesized that sex-determining regions of the genome arose originally as autosomal genes which were incorporated into either a mating type locus or a sex chromosome which then expanded (28). Several types of genes have previ-

ously been found to be present in mating type loci, including those encoding key transcriptional regulators of two classes, the homeodomain proteins (as a pair of interacting proteins of HD1 and HD2 classes), and the HMG domain and  $\alpha$ -box proteins. In a striking example in hemiascomycetes, namely, *Candida albicans*, all four of these genes are present as  $\alpha 1$  and  $\alpha 2$  in the *MTL $\alpha$*  locus and *a1* and *a2* in the *MTLa* locus (17, 35). Recent phylogenetic comparisons have revealed that the HMG domain protein *a2* has been lost in the *Saccharomyces cerevisiae* and *senso stricto* lineages from a common ancestor in which all four genes were present (4).

In the euascomycete lineage, HMG domain and  $\alpha$ -box domain genes are found in the two opposite *MAT* idiomorphs, but the homeodomain genes are not present (Fig. 5). In contrast, in the basidiomycete lineage, paired homeodomain genes are present in the *MAT* locus of all species analyzed thus far (*Ustilago maydis*, *Ustilago hordei*, *Schizophyllum commune*, *Coprinus cinereus*, *C. neoformans*, and *C. gattii*), but HMG domain or  $\alpha$ -box genes are not. A unifying hypothesis is that there are two ancestral fungal sex determinants (one is paired HD1/HD2 homeodomain genes and the other the  $\alpha$ -box and HMG domain pair) and that one or the other or both have been lost from, or acquired by, the *MAT* locus in different lineages (Fig. 5). In some cases, subsequent gene loss events have reduced the number from four to three of these genes, such as in *S. cerevisiae* and related species (4). It is possible that the additional sex-determining genes that are present in the *MAT* locus of some species are instead located elsewhere in the genome in other species and yet remain under the control of the *MAT* locus, such as has been shown for non-*MAT* HMG genes in *U. maydis* (2). Thus, homeodomain genes critical for cell identity and sexual reproduction may remain to be defined in euascomycetes, and HMG domain and  $\alpha$ -box proteins may remain to be defined in members of the phylum *Basidiomycota*.

Whether the ancestral *MAT* locus contained all four sex determinants or whether different lineages coopted one pair or, in some cases, both pairs of determinants remains to be established. One approach to distinguish between these and other models will be to characterize the *MAT* locus from more-divergent fungal lineages, such as zygomycetes. The recent sequencing of the *Phycomyces* genome, with its known sexual cycle and meiotic map, will allow the nature of the *MAT* locus to be defined for comparison with that of ascomycete and basidiomycete fungal species.

Finally, an important consequence of our studies is to enable further insight into the evolution of the *MAT* locus and homothallic and heterothallic sexual cycles in other euascomycete species. In two notable species, *A. nidulans* and *A. fumigatus*, the *MAT* locus is organized in an unusual fashion. *A. nidulans* harbors both *MAT* idiomorphs in the genome of a single isolate at unlinked genomic locales, giving rise to a homothallic sexual cycle. By contrast, in *A. fumigatus*, the two idiomorphs are present at the same genomic locus, and any given isolate harbors only one or the other, but never both. This would be consistent with a heterothallic sexual cycle that remains to be described. Galagan and colleagues have hypothesized that the ancestral organization of the *MAT* locus in aspergilli was one where each isolate contained a locus that included both an HMG domain and an  $\alpha$ -box-encoding gene in a homothallic ancestral species (15). Alternatively, the ancestral species may

have been a heterothallic species from which both heterothallic and homothallic descendants have been derived. Based on our studies, we propose that the ancestor of the *Onygenales* and aspergilli was heterothallic, and if the hypothesized homothallic ancestor to the aspergilli existed, it likely arose after the divergence of the lineages. Furthermore, we hypothesize that the ancestor of euascomycetes and hemiascomycetes may very well have resembled the current structures in *C. albicans* that include two homeodomain genes, paired with either an HMG domain gene or an  $\alpha$ -box gene, but exclude the presence of the *Candida*-acquired *PIK*, *OBP*, and *PAP* genes. Further studies of fungal *MAT* loci will be required to fully understand this complex evolutionary system.

#### ACKNOWLEDGMENTS

This work was supported in part by NIAID AI50113 to J.H., NIAID AI066224 to A.S., and NIAID AI19149 to G.T.C. J.E.S. is supported through the Miller Institute for Basic Research in Science and D.O.I. by an Irvington Institute postdoctoral fellowship.

We thank Alex Idnurm and Anna Floyd for assistance, Meggan Bubnick and George Smulian for discussions and communication of unpublished results, Margareta Andersson for technical assistance, and Dave Des Marais and Anna-Sapfo Malaspinas for discussions regarding positive selection and acknowledge the *H. capsulatum*, *U. reesii*, *C. immitis*, and *C. posadasii* genome projects at the Broad Institute, Washington University at St. Louis, and The Institute for Genomic Research.

#### REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Brefort, T., P. Müller, and R. Kahmann. 2005. The high-mobility-group domain transcription factor Rop1 is a direct regulator of *prf1* in *Ustilago maydis*. *Eukaryot. Cell* **4**:379–391.
- Burt, A., D. A. Carter, G. L. Koenig, T. J. White, and J. W. Taylor. 1996. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* **93**:770–773.
- Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, and K. H. Wolfe. 2004. Evolution of the *MAT* locus and its Ho endonuclease in yeast species. *Proc. Natl. Acad. Sci. USA* **101**:1632–1637.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**:540–552.
- Cole, G. T., J. Xue, K. Seshan, P. Borra, R. Borra, E. Tarcha, R. Schaller, J. J. Yu, and C. Y. Hung. 2006. Virulence mechanisms of *Coccidioides*, p. 363–392. In J. Heitman, S. G. Filler, J. E. Edwards, Jr., and A. P. Mitchell (ed.), *Molecular principles of fungal pathogenesis*. ASM Press, Washington, DC.
- Creevey, C. J., and J. O. McInerney. 2005. Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* **21**:390–392.
- Do, C. B., M. S. Mahabhashyam, M. Brudno, and S. Batzoglou. 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res.* **15**:330–340.
- Fisher, M. C., G. L. Koenig, T. J. White, and J. W. Taylor. 2001. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognised as the non-Californian population of *Coccidioides immitis*. *Mycologia* **94**:73–84.
- Fitzpatrick, D. A., M. E. Logue, J. E. Stajich, and G. Butler. 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* **6**:99.
- Fraser, J. A., S. Diezmann, R. L. Subaran, A. Allen, K. B. Lengeler, F. S. Dietrich, and J. Heitman. 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol.* **2**:2243–2255.
- Fraser, J. A., and J. Heitman. 2004. Evolution of fungal sex chromosomes. *Mol. Microbiol.* **51**:299–306.
- Fraser, J. A., and J. Heitman. 2003. Fungal mating-type loci. *Curr. Biol.* **13**:R792–R795.
- Fraser, J. A., and J. Heitman. 2006. Sex, *MAT*, and the evolution of fungal virulence, p. 13–33. In J. Heitman, S. G. Filler, J. E. Edwards, Jr., and A. P. Mitchell (ed.), *Molecular principles of fungal pathogenesis*. ASM Press, Washington, DC.
- Galagan, J. E., S. E. Calvo, C. Cuomo, L. J. Ma, J. R. Wortman, S. Batzoglou, S. I. Lee, M. Basturkmen, C. C. Spevak, J. Clutterbuck, V. Kapitonov, J. Jurka, C. Scacciochio, M. Farman, J. Butler, S. Purcell, S. Harris, G. H.

- Braus, O. Draht, S. Busch, C. D'Enfert, C. Bouchier, G. H. Goldman, D. Bell-Pedersen, S. Griffiths-Jones, J. H. Doonan, J. Yu, K. Vienken, A. Pain, M. Freitag, E. U. Selker, D. B. Archer, M. A. Penalva, B. R. Oakley, M. Momany, T. Tanaka, T. Kumagai, K. Asai, M. Machida, W. C. Nierman, D. W. Denning, M. Caddick, M. Hynes, M. Paoletti, R. Fischer, B. Miller, P. Dyer, M. S. Sachs, S. A. Osmani, and B. W. Birren. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**:1105–1115.
16. Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
17. Hull, C. M., and A. D. Johnson. 1999. Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. *Science* **285**:1271–1275.
18. James, T. Y., P. Srivilai, U. Kues, and R. Vilgalys. 2006. Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. *Genetics* **172**:1877–1891.
19. Kasuga, T., J. W. Taylor, and T. J. White. 1999. Phylogenetic relationships of varieties and geographical groups of the human pathogenic fungus *Histoplasma capsulatum* Darling. *J. Clin. Microbiol.* **37**:653–663.
20. Koufopanou, V., A. Burt, T. Szaro, and J. W. Taylor. 2001. Gene genealogies, cryptic species, and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (*Ascomycota*, *Onygenales*). *Mol. Biol. Evol.* **18**:1246–1258.
21. Koufopanou, V., A. Burt, and J. W. Taylor. 1997. Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* **94**:5478–5482.
22. Kwon-Chung, K. J. 1981. Virulence of the two mating types of *Emmonsiiella capsulata* and the mating experiments with *Emmonsiiella capsulata* var. *duboisii*, p. 48–56. In C. De Vroey and R. Vanbreuseghem (ed.), *Sexuality and pathogenicity of fungi*. Masson, Paris, France.
23. Kwon-Chung, K. J. 1972. *Emmonsiiella capsulata*: perfect state of *Histoplasma capsulatum*. *Science* **177**:368–369.
24. Kwon-Chung, K. J. 1972. Sexual stage of *Histoplasma capsulatum*. *Science* **175**:326.
25. Kwon-Chung, K. J. 1973. Studies on *Emmonsiiella capsulata*. I. Heterothallism and development of the ascocarp. *Mycologia* **65**:109–121.
26. Kwon-Chung, K. J., R. J. Weeks, and H. W. Larsh. 1974. Studies on *Emmonsiiella capsulata* (*Histoplasma capsulatum*). II. Distribution of the two mating types in 13 endemic states of the United States. *Am. J. Epidemiol.* **99**:44–49.
27. Metznerberg, R. L., and N. L. Glass. 1990. Mating type and mating strategies in *Neurospora*. *Bioessays* **12**:53–59.
28. Ohno, S. 1967. Sex chromosomes and sex-linked genes. Springer-Verlag, New York, NY.
29. Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
30. Pan, S., L. Sigler, and G. T. Cole. 1994. Evidence for a phylogenetic connection between *Coccidioides immitis* and *Uncinocarpus reesii* (*Onygenaceae*). *Microbiology* **140**:1481–1494.
31. Remm, M., C. E. Storm, and E. L. Sonnhammer. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**:1041–1052.
32. Rippon, J. W. 1988. Medical mycology. Saunders, Philadelphia, PA.
33. Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
34. Sil, A., and L. Hwang. 2006. Future of functional genomics of *Histoplasma capsulatum*, p. 611–626. In J. Heitman, S. G. Filler, J. E. Edwards, Jr., and A. P. Mitchell (ed.), *Molecular principles of fungal pathogenesis*. ASM Press, Washington, DC.
35. Tsong, A. E., M. G. Miller, R. M. Raisner, and A. D. Johnson. 2003. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**:389–399.
36. Turgeon, B. G., and O. C. Yoder. 2000. Proposed nomenclature for mating type genes of filamentous ascomycetes. *Fungal Genet. Biol.* **31**:1–5.
37. Woods, J. P. 2002. *Histoplasma capsulatum* molecular genetics, pathogenesis, and responsiveness to its environment. *Fungal Genet. Biol.* **35**:81–97.
38. Woods, J. P. 2006. Molecular determinants of *Histoplasma capsulatum* pathogenesis, p. 321–346. In J. Heitman, S. G. Filler, J. E. Edwards, Jr., and A. P. Mitchell (ed.), *Molecular principles of fungal pathogenesis*. ASM Press, Washington, DC.
39. Worsham, P. L., and W. E. Goldman. 1988. Quantitative plating of *Histoplasma capsulatum* without addition of conditioned medium or siderophores. *J. Med. Vet. Mycol.* **26**:137–143.