Evolution of the Mating Type Locus: Insights Gained from the Dimorphic Primary Fungal Pathogens \textit{Histoplasma capsulatum}, \textit{Coccidioides immitis}, and \textit{Coccidioides posadasii}†‡

James A. Fraser,1 Jason E. Stajich,2 Eric J. Tarcha,3 Garry T. Cole,3 Diane O. Inglis,4 Anita Sii,4 and Joseph Heitman5*

School of Molecular and Microbial Sciences, University of Queensland, Brisbane, QLD 4072, Australia1; Department of Plant and Microbial Biology, University of California, Berkeley, California 947202; Department of Biology, University of Texas at San Antonio, San Antonio, Texas 782493; Department of Microbiology and Immunology, University of California, San Francisco, California 941434; and Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 277105

Received 17 January 2007/Accepted 14 February 2007

Sexual reproduction of fungi is governed by the mating type (\textit{MAT}) locus, a specialized region of the genome encoding key transcriptional regulators that direct regulatory networks to specify cell identity and fate. Knowledge of \textit{MAT} locus structure and evolution has been considerably advanced in recent years as a result of genomic analyses that enable the definition of \textit{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, \textit{Histoplasma capsulatum}, \textit{Coccidioides immitis}, and \textit{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 \textit{H. capsulatum} is not attributable to mutations at the \textit{MAT} locus. These findings also suggest that an extant sexual cycle remains to be defined in both \textit{Coccidioides} species, in accord with population genetic evidence. Based on these \textit{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 \textit{MAT} locus, revealing examples in which flanking genes have been inverted or subsumed and incorporated into an expanding \textit{MAT} locus, allowing us to propose an expanded model for the evolution of the \textit{MAT} locus in the phylum \textit{Ascomycota}.

In the fungal kingdom, sexual reproduction is regulated by a specialized genomic region known as the mating type (\textit{MAT}) locus (12, 13). This important genomic feature has been most extensively studied in the largest of the fungal phyla, the \textit{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 \textit{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 \textit{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 \textit{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 \textit{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 \textit{MAT} locus structures, there is a paucity of information regarding the \textit{Onygenales}, one of the most medically important orders of fungi. The members of the \textit{Onygenales} include a number of dimorphic primary pathogens, including \textit{Histoplasma capsulatum}, \textit{Coccidioides immitis}, and \textit{Coccidioides posadasii}, all capable of causing life-threatening systemic mycoses (6, 38). Although the \textit{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, \textit{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
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 (+/H11002, −/H11001) 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. posadasi*, which are distinct from each other and represent discrete populations undergoing independent sexual recombination (9, 20, 21). While these studies provide evidence that *C. posadasi* 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. posadasi*, and *C. immitis* by combining bioinformatic analyses of genomes, direct sequence confirmation, PCR analysis was performed using the ExTaq polymerase blend (Mirs) with primers given in the following list: JOHE13618, 5′-AGG CAA TAA CAC TGA CGG CAG TAT-3′; JOHE13619, 5′-GCA TTA CAT CGC AGG ACA TGT-3′; JOHE13950, 5′-GTC AGC AAG AAA CCC GGA GTA ATC-3′; JOHE13951, 5′-AAC AAA CTT TCG CAG GCA AGG-3′; JOHE18241, 5′-TGG ACG AGG AAG GAA CCA TTC GTA-3′; JOHE18242, 5′-CTG GTA TTG GTC GTC AGC AAG-3′; JOHE18243, 5′-CAG CCA ATG ACT GGT TCT AAG G-3′; JOHE18244, 5′-TTT ACG GG AAG AGC ATG GTA G-3′; JOHE18245, 5′-AGG AAA CGA TGT CTC-3′; JOHE18246, 5′-TAC AGG AAG GTA ATC TGG G-3′; JOHE18247, 5′-AAC AAG CAA CGA ACA AGG CTT-3′; and JOHE18248, 5′-AGG TGA GAG GAC AAG AAC AAT GAG-3′.

**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 rash-23; *C. immitis* strains RS and H538.4, and *C. posadasi* 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 were grown on soil extract agar at room temperature and on Histoplasma macrophage medium plates at 37°C under 5% CO₂ (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% CO₂ 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 posadasi* 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, the end sequences 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 DOTTRIP 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. posadasi* MAT1-2 idiomorph. Primers JOHE18241 to JOHE18248 were used for MAT diagnostics. PCR analysis was performed using the ExTaQ polymerase blend (Mirs) with primers given in the following list: JOHE13618, 5′-AGG CAA TAA CAC TGA CGG CAG TAT-3′; JOHE13619, 5′-GCA TTA CAT CGC AGG ACA TGT-3′; JOHE13950, 5′-GTC AGC AAG AAA CCC GGA GTA ATC-3′; JOHE13951, 5′-AAC AAA CTT TCG CAG GCA AGG-3′; JOHE18241, 5′-TGG ACG AGG AAG GAA CCA TTC GTA-3′; JOHE18242, 5′-CTG GTA TTG GTC GTC AGC AAG-3′; JOHE18243, 5′-CAG CCA ATG ACT GGT TCT AAG G-3′; JOHE18244, 5′-TTT ACG GG AAG AGC ATG GTA G-3′; JOHE18245, 5′-AGG AAA CGA TGT CTC-3′; JOHE18246, 5′-TAC AGG AAG GTA ATC TGG G-3′; JOHE18247, 5′-AAC AAG CAA CGA ACA AGG CTT-3′; and JOHE18248, 5′-AGG TGA GAG GAC AAG AAC AAT GAG-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 passage. The second G217B isolate was obtained directly from the American Type Culture Collection (ATCC 26032) and presumably has not been passed 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 AlphaCol 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 (α-box)-specific PCR employed primers JOHE18245 and JOHE18246 while MAT1-2 (HMG domain)-specific PCR employed primers JOHE18247 and JOHE18248. For *Coccidioides*, MAT1-1 (α-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 α-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 SL2 (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 α-box locus is associated with SL2, 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 α-box transcription factor homolog. The structure of this MAT1 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 MAT1 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 SL2 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 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 MAT1 idiomorph of ATCC 22636 is 5,418 bp in length and encodes an α-box protein. In contrast, the MAT1 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 α-box and HMG domain MAT1 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).
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 revealed 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 α-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 α-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 α-box idiomorph as we had hypothesized. Employing this new sequence, we designed Coccidioides-specific 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 α-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 reessii (30), which encodes a MAT1-1 α-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 α-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
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 compar-
ing 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 en-
donuclease 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 transi-
tional 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 His-
toplasma, 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-
cans, where, in addition to the mating type transcription fac-
tors, 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 incor-
porated 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.
neofoms 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 repro-
duction). 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 jux-
taposition 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, His-
toplasma capsulatum, Coccidioides immitis, and Coccidioides
posadasii, all of which are primary pathogens of humans. A
sexual cycle has been previously described only for H. capsu-
latum (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 α-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 het-
erothallic 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 ge-
nome, and thus, all three are molecularly distinguished from
homothallic species that have been defined in the fungal king-
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 α-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 α-box transcription factor, similar to those of other euascomycete 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 α-box proteins. In a striking example in hemiascomycetes, namely, *Candida albicans*, all four of these genes are present as α1 and α2 in the *MTLa* locus and α1 and α2 in the *MTLa* locus (17, 35). Recent phylogenetic comparisons have revealed that the HMG domain protein α2 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 α-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*, *Schizosaccharomyces commune*, *Coprinus cinereus*, *C. neoformans*, and *C. gattii*), but HMG domain or α-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 α-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 α-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 loci, 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 α-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 α-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 AI91194 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 Idrurum 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 Malaspina for discussions regarding positive selection and acknowledge the *H. capsulatum*, *U. reesti*, *C. immitis*, and *C. posadasi* genome projects at the Broad Institute, Washington University at St. Louis, and The Institute for Genomic Research.

**REFERENCES**


