

# Molecular Evolution of the Fungi: Human Pathogens<sup>1</sup>

Barbara H. Bowman,\* John W. Taylor,† and Thomas J. White\*

\*Roche Molecular Systems; and †Department of Plant Biology, University of California, Berkeley

The morphological, ecological, and clinical diversity among ascomycete fungi that are pathogenic to humans suggest that the potential for pathogenicity may have arisen multiple times within these higher fungi. We have obtained 18S ribosomal DNA sequences from a diverse group of human pathogenic fungi in order to determine their evolutionary origins. The fungi studied include a skin pathogen that is confined to humans (*Trichophyton rubrum*) and three systemic, facultative parasites that cause histoplasmosis (*Histoplasma capsulatum*), blastomycosis (*Blastomyces dermatitidis*) and coccidioidomycosis (*Coccidioides immitis*) in humans and other higher animals. Also included in our analysis are representatives of non-pathogenic fungi, as well as two opportunistic pathogens, *Pneumocystis carinii* and *Candida albicans*, that cause severe disease in immunocompromised individuals, especially those with AIDS. Two of the fungi we sequenced, *T. rubrum* and *C. immitis*, are limited to asexual modes of reproduction and therefore lack the sexual structures that are most useful for evolutionary comparison as well as being essential for classification among the higher fungi. *Coccidioides immitis* is particularly problematic owing to its contradictory and confusing asexual morphologies, which have caused it to be placed in three fungal classes and the protista. Our analysis shows that the specialized, superficial parasite and the systemic, facultative parasites, including *C. immitis*, are closely related ascomycetes, which clearly demonstrates the power of molecular characters to compensate for missing or confusing reproductive morphology. Analysis also shows that the opportunistic pathogens are more distantly related, with the likely explanation that pathogenicity has arisen more than once within the Ascomycetes.

## Introduction

Fungi are the cause of several life-threatening diseases in humans, including pneumonia, septicemia, and skin and systemic disease (Rippon 1988, pp. 2 and 154–797). Immunocompromised individuals are especially susceptible (Galgiani and Ampel 1990; Wheat et al. 1990). Human pathogenic fungi are a diverse group morphologically, ecologically, and clinically (Rippon 1988, pp. 4 and 154–797), which suggests that the potential for pathogenicity may have arisen multiple times. To examine the origins of pathogenicity, it is necessary to determine the evolutionary relationships of pathogens to each other and to other fungi.

Fungal phylogenetics has been well served by morphological comparison, particularly of sexual reproductive structures. Where sexual morphology is lacking, evolutionary analysis has been more difficult. Morphological analysis of asexual fungi is

1. Key words: molecular evolution, 18S ribosomal RNA, pathogenic fungi, ascomycete, medical mycology.

Address for correspondence and reprints: Barbara H. Bowman, Roche Molecular Systems, Inc., 1145 Atlantic Avenue, Alameda, California 94501.

*Mol. Biol. Evol.* 9(5):893–904, 1992.  
© 1992 by The University of Chicago. All rights reserved.  
0737-4038/92/0905-0010\$02.00

made even more difficult when the characters are confusing or in conflict, as is the case with one pathogen in this study, *Coccidioides immitis*. DNA sequence analysis offers a straightforward means of resolving evolutionary questions where morphological characters are absent or contradictory.

We have obtained 18S rDNA sequences for four fungal pathogens, which vary in their ecology, morphology, and clinical presentation. The fungus *Histoplasma capsulatum*, present in soil associated with bird or bat excrement, is found throughout the world and is endemic to the Mississippi and Ohio River Valleys (Rippon 1988, p. 383). *Coccidioides immitis*, which dwells in soil, is found in the subtropical New World and is endemic to the San Joaquin Valley and the Sonoran desert of western North America (Galgiani and Wack 1988). The diseases caused by these two fungi elicit detectable immune responses (Ampel et al. 1989; Pappagianis and Zimmer 1990) and are usually benign but will occasionally cause disseminated infections requiring specific treatment to avoid fatality. *Blastomyces dermatitidis* causes nonspecific symptoms, is difficult to detect immunologically (Sarosi et al. 1988; Jorday et al. 1990), and is among the last microbes to be considered in differential diagnoses. Detection of *B. dermatitidis* in nature is also difficult, although the fungus appears to inhabit the same area in North America as does *H. capsulatum* (Klein et al. 1986; Rippon 1988, p. 477). *Trichophyton rubrum* is the cosmopolitan, anthropophylic, and typically chronic infection of the smooth skin and nails (Rippon 1988, p. 218).

Many higher fungi have both sexual and asexual (clonal) modes of reproduction, often with strikingly different morphologies. The sexual and asexual reproductive states of these fungi may be named and classified independently, though genetically they are the same organism and share the same evolutionary history. The names *H. capsulatum* and *B. dermatitidis* apply only to the asexual states of these pathogenic fungi, which are the spore states encountered by clinicians. Under their asexual names, these fungi are classified in the form division Deuteromycota (=Fungi Imperfecti). Deuteromycota is termed a "form division" because the classification of fungi in this division is not recognized based on phylogeny. These same fungi are known to mycologists by the names *Ajellomyces capsulatus* and *A. dermatitidis*, which refer to both their sexual and asexual states (McGinnis and Katz 1979). Their classification in the class Ascomycetes is based on their sexual reproductive morphology.

Many fungi apparently lack sexual reproduction and, with it, the morphological features that facilitate evolutionary comparison and allow formal classification among the higher fungal classes (Greuter et al. 1988, p. 60). *Trichophyton rubrum* and *C. immitis* are examples of such asexual fungi. Asexual morphology may suggest evolutionary affinities, as in the case of *T. rubrum*, where asexual spore states closely resemble those of other ascomycete skin pathogens (Rippon 1988, p. 256). However, where asexual morphology is confusing, as in the case of *C. immitis*, phylogenetic placement has been controversial (Stevens 1980, p. 44; Rippon 1988, p. 469). *Coccidioides immitis* has two asexual spore types, one formed in the environment and the other formed in animal tissue. The environmental spores support an affiliation with Ascomycota (Sigler and Carmichael 1976). The spherules and endospores formed in tissue have been used to place *C. immitis* in the protista (Rixford and Gilchrist 1896; Rippon 1988, p. 433) as well as in three fungal divisions—the Ascomycota (Ophüls 1905), Chytridiomycota (Ciferri and Redaelli 1936), and Zygomycota (Baker et al. 1943)—but also to exclude classification in the Chytridiomycota and the Zygomycota (Cole and Sun 1985). DNA sequence analysis should provide the characters that will

lead to a unified, phylogenetically based classification system that embraces both sexual and asexual fungi.

## Material and Methods

*Histoplasma capsulatum* (ATCC 11408) *Blastomyces dermatitidis* (ATCC 26199) and *Coccidioides immitis* (ATCC 28868) were obtained from the American Type Culture Collection (ATCC) and grown in culture. DNAs were extracted according to a method described elsewhere (Lee and Taylor 1990), using BL-3 precautions. *Trichophyton rubrum* DNA (IFO 9185) was a gift of G. Apodaca and J. Sakanari. Nearly full-length 18S rDNAs were amplified using the polymerase chain reaction (PCR) with primers NS1 (White et al. 1990) and NS24 (Bowman et al. 1992), one of which was biotinylated in each of two reciprocal reactions. Single-stranded DNA for direct sequencing was obtained using the streptavidin agarose capture method of Mitchell and Merrill (1989), as revised in our laboratory (Bowman et al. 1992; Gargas and Taylor 1992). This procedure yielded sufficient ssDNA for four or five direct sequencing reactions from each 100- $\mu$ l PCR amplification. Single-stranded DNAs were sequenced with multiple internal primers by using a standard Sequenase protocol (U.S. Biochemical), except that labeling mix was diluted 1/20. For the four sequences presented data were obtained for both strands of the DNA at 92%–96% of the positions; for the remaining 4%–8% of positions, sequences presented are based on information from only one strand. Sequences from the primer regions are not included in figures or analysis.

Sequences were aligned by hand, using the Eyeball Sequence Editor (ESEE 1.09) (Cabot and Beckenbach 1989). Positions that were not alignable on the basis of primary structure were excluded from evolutionary analyses. Parsimony trees were drawn using the exhaustive search option on PAUP 3.0q (Swofford 1990), on the basis of 16 phylogenetically informative sequence positions (PI sites). Of these, 144 PI sites were from sequence positions alignable in all species, and 23 were from sequence positions alignable in subsets of the species. (Unaligned sequences at these positions were coded as “no data.”) One PI site contained a T in the four new pathogen sequences and a single-base deletion in all the other fungi; no other PI site contained or was bounded by a deletion. All positions in figure 1 were included in parsimony analyses. Bootstrap analyses used the branch-and-bound option of PAUP 3.0q (Swofford 1990), with 1,000 repeat samplings of the data. The distance tree was drawn using NJTREE (Jin and Ferguson 1990), on the basis of 335 variable sequence positions within the 1,565 positions alignable in all 10 species. Positions where one or more species contained length mutation were not included in distance analyses. Pairwise substitutions were uncorrected, the maximum difference being 190, or 12.1%, between *Neurospora crassa* and *Chytridium confervae* (table 1).

## Results

For each pathogen, 1,713 bases of 18S rDNA sequence were obtained, spanning the length between the amplification primers and thus excluding an estimated 38 and 49 bases, respectively, at the 5' and 3' ends of the 18S molecule. Our sequence for *Coccidioides immitis* differs from the unpublished sequence submitted to GenBank (COIDA), which has a T at our position 483 (C) and a single C (rather than two) at our positions 1134–1135. Our sequence has been reverified on original gels. Our *Blastomyces dermatitidis* sequence differs from the unpublished sequence submitted to GenBank (BLODA) in that we have resolved two bases that were ambiguous.

Hc AAAGATTAAGCCATGCATGTCTAAGTATAAGCAATCTATACGGTGAACCTGCGAATGGCTCATTAATCAGTTATCGTTTATTGTAGTAGCTTACTAC 100  
 Bd .....T.G.....  
 Ci .....AT.G.....  
 Tr .....  
 -----  
 Hc TTGGATACCCGTGGTAATCTAGAGCTAATACATGCTGAAAGCCTCGACTTCGGAAGGGGTGATTATTAGATAAAAAACCAATGCCCTTCGGGGCTCC 200  
 Bd .....A.A.....  
 Ci .....A.A.....T  
 Tr .....A.A.....C.....T  
 -----  
 Hc TTGGTGATTCATAAATACTTCGAATCGCATGGCTTGCSCGGCGGATGGTTCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCT 300  
 Bd .....A.....T.....  
 Ci .....C.....  
 Tr .....C.....  
 -----  
 Hc ACCATGGTGGCAACGGGTAAACGGGAATTAGGGTTCGATTCCGGAGAGGGACCTGAGAGACGGCTACCACATCTAAGGAAGGCAGCAGGCCGCCAAAT 400  
 Bd .....A.....C.....  
 Ci .....A.....C.....  
 Tr .....A.....C.....  
 -----  
 Hc ACCCAATCCCAGATACGGGGAGTGTGACATAAATACTGATACAGGGCTCTTTCGGGTCTTGTAAATGGCAATGAGTACAATCTAAATCCCTTAACGAGG 500  
 Bd .....A.C.....  
 Ci .....A.....T.....C.....  
 Tr .....A.....T.....C.....  
 -----  
 Hc AACAAATGGAGGGCAAGTCTGGTCCAGCAGCCGGGTAATCCAGCTCCAATAGCGTATATAAAAGTTGTGCAGTTAAAAAGCTCGTAGTTGAACCTT 600  
 Bd .....  
 Ci .....  
 Tr .....  
 -----  
 Hc GGCTCTGGCTGGCCGGTCCGCCTCACCCTGCACTGGTCCGGCCGACCTTTCCTCTGGGGAGCCTCAGCCCTCAGTGGCTGTGGGGGAACCAAGG 700  
 Bd .....A.....C.....T.....  
 Ci .....T.G.G.....T.....T.....A.....CT.....T.....A.....  
 Tr .....C.....T.....G.....T.....G.....T.....C.....A.....C.....C.....C.....  
 -----  
 Hc ACTTTTACTGTGAAAAAATTAGAGTGTTCAAAGCAGGCCCTTTCCTCGGATACATTAGCATGGAATAATAGAATAGGACGCTGTGGTTCTATTTTGTGGTT 800  
 Bd .....  
 Ci .....G.....  
 Tr .....  
 -----  
 Hc TCTAGGACCCCGTAAATGATTAAATAGGGATGGTGGGGCGTCACTATTCCGGCTGTGAGAGTCAAATCTTGGATTGCTGAAGACTAACTACTGGCAA 900  
 Bd .....A.....  
 Ci .....A.....C.....  
 Tr .....A.....C.....  
 -----  
 Hc AGCATTCCGCAAGGATGTTTTATTAAATCAGTGAACGAAAGTTAGGGATCGAAGCAGATCAGATACCGTCTGTAGTCTTAACCAATAAACTATGCCGACTA 1000  
 Bd .....  
 Ci .....  
 Tr .....  
 -----  
 Hc GGGATCGGACGGGGTCTTATGATGACCCCTTCGGCACCTTACGAGAAATCAAAGTTTTGGGTTCTGGGGGAGTATGGTCCGCAAGGCTGAAACTAAA 1100  
 Bd .....CAA...TGA..A..  
 Ci .....TC.T.T..  
 Tr .....  
 -----  
 Hc GAAATGACGGAAGGGCACCACAGGCGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACTACCAGGTCCAGACAAAATAAGGATTGACAGAT 1200  
 Bd .....C.....  
 Ci .....C.....  
 Tr .....  
 -----  
 Hc TGAGAGCTCTTCTTGATCTTTGGATGGTGTGATGCCGTTCTTAGTGTGGAGTATTGTCTGCTTAATTTGGCATAACGAACGAGACCTTAACC 1300  
 Bd .....  
 Ci .....  
 Tr .....  
 -----  
 Hc TGCTAAATAGCCGACCCGGCTCTGGCCGGCTGGCTTCTTAGAGGGACTATCGGCTCAAGCCGATGGAAGTTTGAGGCAATAACAGGCTGTGTATGCC 1400  
 Bd .....G.T...T..  
 Ci .....A...T..  
 Tr .....GT.G...C.C..  
 -----  
 Hc CTTAGATGTTCTGGCCGCACGCGCTACACTGACAGGGCCAGCGTACATCACCTTGACCGAGAGGTCTGGGTAATCTTGTTAAACCCTGTCTGTCT 1500  
 Bd .....G.....  
 Ci .....G.....  
 Tr .....C.....G.....C.C.....  
 -----  
 Hc GGGGATAGAGCATTGCAATATTGCTCTTCAACGAGGAATGCCTAGTAGGCACAAGTCATCAGTTGTGCCGATTACGTCCCTGCCCTTTGTACACACCG 1600  
 Bd .....  
 Ci .....  
 Tr .....  
 -----  
 Hc CCCGTCGCTACTACCGATTGAATGGCTCAGTAGGCCCTTCGACTGGCTCAGGGGGTTGGCAAACGCCGCCAGAGCCGGAAGTTGGTCAAACCTGGT 1700  
 Bd .....A.....A.....  
 Ci .....C.....A.....A.....G.....  
 Tr .....C.....A.....A.....G.....C.....  
 -----  
 Hc CATTTAGAGGAAG 1713  
 Bd .....  
 Ci .....  
 Tr .....

## Phylogeny

The evolutionary tree in figure 2A was constructed on the basis of the principle of maximum parsimony (Camin and Sokal 1965), using PAUP 3.0q (Swofford 1990). This tree was the single most parsimonious solution, with a length of 330, on the basis of 167 phylogenetically informative sequence positions. Because of autapomorphies (unique substitutions) within the phylogenetically informative sites, the minimum possible length would be 222. There are an additional 211 uniquely derived characters within the sequence positions alignable in all 10 taxa. This tree has a consistency index of 0.673, with a retention index of 0.684 (Farris 1989). The flagellated fungus *Chytridium confervae* serves as an outgroup to the higher fungal classes Ascomycetes and Basidiomycetes (Bowman et al. 1992). Branch lengths in figure 2A are not proportional to the inferred number of substitutions, because, for 23 of the sequence positions in the analysis, one or more sequences were unalignable and could not be included. This would cause the number of substitutions on these branches to be underestimated.

The most parsimonious tree indicates that the systemic and skin pathogens, including the two asexual species, are closest relatives. The asexual fungi *Coccidioides immitis* and *Trichophyton rubrum* can now confidently be placed in the class Ascomycetes, represented in our study by the bread mold *Neurospora crassa*. Among the facultative pathogens, *Histoplasma capsulatum* and *B. dermatitidis* are closest relatives; these are the two species that share the single genus name "Ajellomyces" in their sexual states. The specialized skin pathogen *T. rubrum* branches outside the three facultative pathogens.

The opportunistic infectious agents *Candida albicans* and *Pneumocystis carinii*, however, are not closely related to the other pathogens. *Candida* represents an early branch within the Ascomycetes, and its closest relative among the species studied is the yeast *Saccharomyces cerevisiae*. A similar affinity was reported by Hendriks et al. (1989) and Barns et al. (1991). An even earlier branch is represented by the AIDS opportunistic fungus *P. carinii* (Edman et al. 1988), which is not closely related to *Candida albicans* or to the other pathogens in the present study.

The exhaustive search option of PAUP 3.0q includes a measure of the skewness of the distribution of all tree lengths, to help determine whether there is phylogenetic information or mainly noise in the data set. If the statistic *gl* is strongly negative, indicating a left skewing of the tree distribution (a small number of trees are significantly better than the rest), then there is a high degree of confidence that the true tree falls near the most parsimonious tree in the distribution (Hillis 1991; Hillis and Huelsenbeck 1992). Our tree length distribution was highly left skewed, with a *gl* of  $-0.64$ , greatly exceeding the critical value for significance at the 99% level ( $-0.33$  for 10 taxa with 100 PI sites,  $-0.27$  with 250 PI sites) (Hillis 1991). Such a distribution demonstrates a highly structured data set with a large degree of internal consistency that is not offset by noise.

FIG. 1.—Aligned 18S rDNA sequences from four human pathogenic fungi: *Histoplasma capsulatum* (Hc), *Blastomyces dermatitidis* (Bd), *Coccidioides immitis* (Ci), and *Trichophyton rubrum* (Tr). A dot (.) indicates a base that is the same as in the reference (i.e., top) sequence. Sequences were aligned by hand by using ESEE 1.09 (Cabot and Beckenbach 1989). The four pathogen sequences obtained are unambiguously alignable at all 1,713 positions. The 1,565 positions used in distance analyses are underlined. GenBank accession numbers for the sequences are as follows: *H. capsulatum*, X58572; *B. dermatitidis*, X59420; *C. immitis*, X58571; and *T. rubrum*, X58570.

**Table 1****Number of Pairwise Differences in 1,565 Aligned Sequence Positions**

	<i>Blastomyces dermatitis</i>	<i>Coccidioides immitis</i>	<i>Trichophyton rubrum</i>	<i>Neurospora crassa</i>	<i>Candida albicans</i>	<i>Saccharomyces cerevisiae</i>	<i>Pneumocystis carinii</i>	<i>Spongipellis unicolor</i>	<i>Chytridium confervae</i>
<i>Histoplasma capsulatum</i> . . . . .	6	19	28	108	102	108	124	143	165
<i>B. dermatitidis</i> . . . . .		15	23	103	98	104	119	142	163
<i>C. immitis</i> . . . . .			24	102	102	105	121	142	167
<i>T. rubrum</i> . . . . .				98	106	115	125	145	162
<i>N. crassa</i> . . . . .					131	136	142	168	190
<i>C. albicans</i> . . . . .						52	108	141	152
<i>S. cerevisiae</i> . . . . .							117	146	158
<i>P. carinii</i> . . . . .								132	153
<i>S. unicolor</i> . . . . .									165

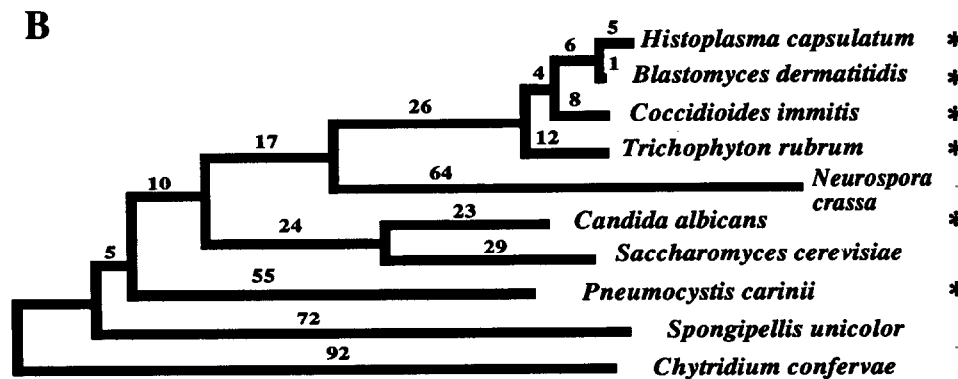
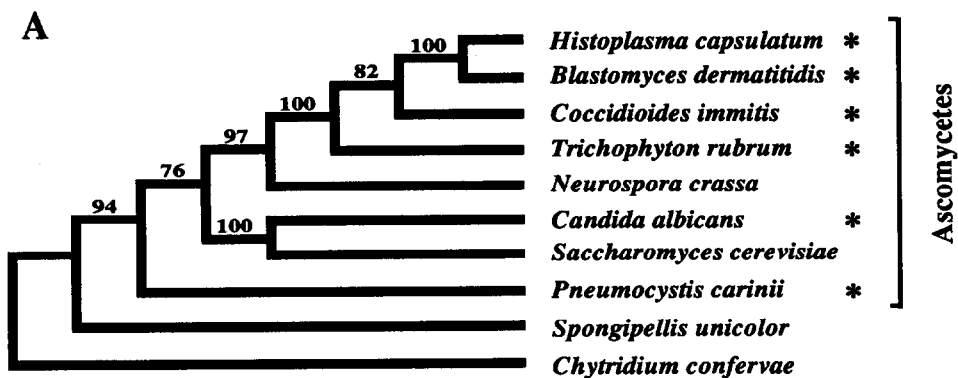


FIG. 2.—Parsimony and distance trees showing the evolutionary relationships among six human pathogens and the higher fungi, using a chytridiomycete (water mold) as the outgroup. Pathogens are marked with an asterisk (\*). The four pathogen sequences in fig. 1 were aligned with the 18S rDNA sequences from one basidiomycete (*Spongipellis unicolor*, GenBank M59760) two ascomycetes (*Neurospora crassa*, NEURRNAS; and *Saccharomyces cerevisiae*, YSCRGEA), one chytridiomycete (*Chytridium confervae*, M59758), and two opportunistic pathogens (*Candida albicans*, EMBL CAL16S; and *Pneumocystis carinii*, GenBank PMC16SRR1). The 10 sequences were alignable at the 1,565 positions underlined in fig. 1. (Alignment will be supplied on request.) Additional positions were included in parsimony calculations when they were alignable for a subset of the sequences. All positions in fig. 1 were included in parsimony analyses. A, Tree based on the principle of maximum parsimony (Camin and Sokal 1965) employing the computer program PAUP 3.0q (phylogenetic analysis using parsimony) (Swofford 1990). Numbers above the internal branches were determined using the bootstrapping option of PAUP; these indicate the robustness of the branching order, by showing the percentage (rounded to the nearest whole number) of times, that each group was monophyletic in 1,000 repeat subsamplings of the data. Numbers  $\geq 95$  define a branch strongly supported by the data. The tree is based on 134 phylogenetically informative positions in the sequence areas alignable in 18S rDNAs from all 10 species, with the addition of 23 phylogenetically informative positions alignable only in a subset of the species. These include positions 36, 200, 455, and 1017, omitting species with insertions or deletions; at other positions the species shown in parentheses were omitted: 621 (*Saccharomyces cerevisiae* and *Chytridium confervae*); 627 (*Candida albicans*, *Saccharomyces cerevisiae*, *Pneumocystis carinii*, and *Candida albicans*); 633 (*Pneumocystis carinii*); 639, 642, 645, 647, 648, and 650 (*Saccharomyces cerevisiae*); 662 and 665 (*Candida albicans*, *Saccharomyces cerevisiae*, *Spongipellis unicolor*, and *Chytridium confervae*); 673, 678, 684, 689, and 670 (*Candida albicans* and *Saccharomyces cerevisiae*); and 1315, 1317, and 1323 (*Candida albicans*, *Saccharomyces cerevisiae*, *Pneumocystis carinii*, *Spongipellis unicolor*, and *Chytridium confervae*). This tree shows the order of branching only. Because the above 23 phylogenetically informative sites could not be coded for the taxa whose sequences were unalignable, the number of substitutions per branch would not accurately reflect the relative branch lengths. B, Tree based on the genetic distances among species and drawn using the neighbor-joining (Saitou and Nei 1987) algorithm, employing the computer program NJTREE (Jin and Ferguson 1990), based on the pairwise substitutions in table 1. Of the 1,565 sequence positions alignable in all 10 taxa, 335 showed at least one substitution. The figure on each branch shows the number of DNA substitutions inferred to have been fixed on that branch. This tree has the same branching order as the tree in panel A.

Examining trees near the most parsimonious tree in length may show which branching orders are most vulnerable to change. Following the 330-step most parsimonious tree shown in figure 1 are two trees of length 333; these differ from the most parsimonious tree in reversing the branching order of *T. rubrum* and *Coccidioides immitis* or in making them sister groups. In a single tree of length 334 *P. carinii* forms the sister group to the *Candida/Saccharomyces* branch, and, in the next tree (336) *Candida/Saccharomyces* branches off before *P. carinii*. The relationships within the group of new pathogen sequences are defined by only 13 PI sites. Of these, 11 support the unity of *H. capsulatum* and *B. dermatitidis*. It costs eight additional mutations to place *Coccidioides immitis* as the sister group to *H. capsulatum* or to make the basidiomycete *Spongipellis unicolor* branch inside *P. carinii*.

To determine whether the branching order in figure 2 is strongly supported, we evaluated the tree by using the bootstrapping algorithm of PAUP 3.0q (Swofford 1990). The percentage of trees in which a given branch was monophyletic is shown by the number above the tree's stem in figure 2A. Branches that are monophyletic in  $\geq 95\%$  of the 1,000 trials are considered strongly supported, so long as the number of informative positions is greater than three (Felsenstein 1985). For example, 11 positions support—and only two oppose—the association of *H. capsulatum* and *B. dermatitidis*, to the exclusion of *Coccidioides immitis* and *T. rubrum*. The branch uniting the three systemic pathogens and the skin pathogen appeared in all 1,000 bootstrap replications. The close relationship of *H. capsulatum* and *B. dermatitidis* was also strongly supported (998 replications), though the specific branching order of *Coccidioides immitis* and *T. rubrum* was not (824 replications). Also strongly supported are both the placement of all four of these pathogens in the class Ascomycetes (as specific relatives of the bread mold *N. crassa*; 966 replications) and the relationship between the yeast *Saccharomyces cerevisiae* and the yeastlike pathogen *Candida albicans* (999 replications). The association of *P. carinii* with the ascomycetes was supported by 937 of the 1,000 replications. This is slightly short of the level needed for confidence in the association. Because it seems likely that *P. carinii* will be found to belong in the Ascomycetes, we assume this for simplicity in this discussion. This assumption does not affect our conclusions.

For a second verification of the branching order, we used a genetic distance-derived "neighbor-joining" algorithm (Saitou and Nei 1987) to redraw the tree (fig. 2B), on the basis of 335 variable positions among the 1565 DNA sequence positions unambiguously alignable in all 10 species. The pairwise distance matrix is given in table 1. The branching order of this tree was identical to that determined by parsimony. Branch lengths in figure 2B illustrate the relative genetic distances among the fungi.

## Molecular Evolution

Base compositions do not vary appreciably among the 18S sequences of these 10 fungi. Each of the four new pathogen sequences has a G+C content of 49%. Within the comparable region of sequence, the remaining ascomycetes had slightly lower G+C content: *N. crassa*, 48%; *Candida albicans*, 46%; and *Saccharomyces cerevisiae* and *P. carinii*, each 45%. The basidiomycete *Spongipellis unicolor* was within this range, with a G+C content of 47%, and the chytrid *Chytridium confervae* had the lowest, at 44%.

Of the 1,565 sequence positions aligned in all 10 species, 335, or 21%, contained one or more substitutions. Of these, 54 contained a substitution only in the branch leading to the outgroup; thus, within the higher fungi, only 281 positions (18%) were



variable. Among the four newly sequenced pathogens, only 22 positions (1.4%) contained substitutions.

In a tree-based analysis, transition substitutions outnumbered transversions by a factor of 1.9. Within the higher fungi (when the outgroup *Chytridium confervae* is excluded), the ratio of transitions to transversions increased to 2.3, and within the four newly sequenced pathogens, transitions were 3.7 times as numerous as transversions. This would suggest that transition mutations have outnumbered transversions but have been obscured increasingly over time by additional transitions, while transversions are less likely to have been obscured by subsequent transversions (Brown et al. 1982).

The two closest relatives, *H. capsulatum* and *B. dermatitidis*, differ at only 12 positions within the 1,713 bases sequenced. They differ from *Coccidioides immitis* by 35 and 33 substitutions, respectively, and these three differ from *T. rubrum* by 47–50 substitutions. There are no length mutations among these four sequences. The deepest divergence among the pathogens in this study is between *P. carinii* and the remaining Ascomycetes. *Pneumocystis carinii* differs from *H. capsulatum*, for example, by 186 substitutions and 11 insertions/deletions (indels). To put these differences into perspective, the *H. capsulatum*/*B. dermatitidis* distance is roughly comparable to that between the 18S sequences of two species of the red alga *Gracilaria* (18 substitutions within 1,771 aligned sequence positions), between human and mouse (13 substitutions and 5 indels within 1,871 aligned positions), or between rat and mouse (7 substitutions and 5 indels within 1,860 positions). The greatest number of substitutions among the four systemic and skin pathogen sequences is 50, about  $\frac{2}{3}$  of the number between the human sequence and that of the frog *Xenopus laevis* (75 substitutions and 8 indels within 1,813 alignable positions). The maximum number of substitutions among the new pathogens is also less than that observed in a deep divergence between two dicotyledonous plants, represented by *Arabidopsis* and soybean (62 substitutions and 9 indels within 1,810 positions), or between rice, a monocot, and soybean (86 substitutions and 9 indels in 1,814 positions). The deepest pathogen divergence observed here, that between *P. carinii* and the rest of the Ascomycetes, has resulted in about half as many substitutions as are observed between a green plant (soybean) and a red algal protist (*Gracilaria*), which is an outgroup to the plant/fungal divergence (357 substitutions and 27 indels within 1,763 aligned positions).

## Discussion

Our analysis of these 18S rDNA sequences shows conclusively that the four systemic and skin pathogens are close relatives, compared with *Neurospora*, yeast, and the opportunistic pathogens. In addition, their relationship to the bread mold *N. crassa* to the exclusion of the two opportunistic pathogens, is strongly supported. Equally strong is the relationship of the opportunistic pathogen *Candida* to the common yeast *Saccharomyces*, to the exclusion of *Pneumocystis*. Thus the parasitic habit must have been acquired independently at least three times among the ascomycete pathogens. Formally, multiple losses—or gains and losses—could also account for this pattern. However, the common ancestor of the Ascomycetes certainly predates the origin of humans and probably predates the origin of mammals, so this hypothesis is unlikely. The four closely related pathogens could be descendants of a single pathogenic lineage, which could have originated during the existence of mammals. In order to resolve this question, we will compare their sequences with those of possible close fungal relatives that are not pathogens.

This study also demonstrates the ability of DNA sequence analysis to place asexual fungi among their sexual relatives, with a high degree of confidence. *Coccidioides immitis* is certainly an ascomycete, as suggested by its asexual arthrospores (arthroconidia) (Sigler and Carmichael 1976); this molecular phylogeny should contribute to resolving the longstanding systematic controversy stimulated by the unusual endospore-spherule mode of reproduction seen in *Coccidioides immitis* within its human and animal hosts. (Stevens 1980, p. 44).

Infections by *Coccidioides immitis* and *Histoplasma capsulatum*, as well as by *Candida albicans* and *P. carinii*, are likely to increase as the number of immunocompromised patients rises (Galgiani and Ampel 1990; Kurtin et al. 1990). At present, unequivocal identification of human mycoses still depends on time-consuming examination of fungal cultures (Keath et al. 1989; Sugar 1989; Antoniskis et al. 1990; Jordan et al. 1990). The sequence variability exhibited by these fungi can form the basis for rapid and specific detection and identification using DNA amplification and oligonucleotide probes.

### Acknowledgments

We thank M. Berbee, E. Reiss, D. Dodge, M. McGinnis, and E. Zimmer for helpful comments on the manuscript; T. Szaro for technical advice; and G. Apodaca and J. Sakanari for providing *Trichophyton rubrum* DNA. This work was supported by NIH grant RO1-AI28545.

### LITERATURE CITED

- AMPEL, N. M., M. A. WIEDEN, and J. N. GALGIANI. 1989. Coccidioidomycosis: clinical update. *Rev. Infect. Dis.* **11**:897-911.
- ANTONISKIS, D., R. A. LARSON, B. AKIL, M. U. RARICK, and J. M. LEEDOM. 1990. Seronegative disseminated coccidioidomycosis in patients with HIV infection. *AIDS* **4**:691-693.
- BAKER, E. E., E. M. MRAK, and C. E. SMITH. 1943. The morphology, taxonomy, and distribution of *Coccidioides immitis* rixford and gilchrist 1896. *Farlowia* **1**:199-244.
- BARNES, S. M., D. J. LANE, M. L. SOGIN, C. BIBEAU, and W. G. WEISBURG. 1991. Evolutionary relationships among Pathogenic *Candida* species and relatives. *J. Bacteriol.* **173**:2250-2255.
- BOWMAN, B. H., J. W. TAYLOR, A. G. BROWNLEE, J. LEE, S.-D. LU, and T. J. WHITE. 1992. Molecular evolution of the fungi: relationship of the Basidiomycetes, Ascomycetes, and Chytridiomycetes. *Mol. Biol. Evol.* **9**:285-296.
- BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial sequences of Primates: tempo and mode of evolution. *J. Mol. Evol.* **18**:225-239.
- CABOT, E. L., and A. T. BECKENBACH. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Comput. Appl. Biosci.* **5**:233-234.
- CAMIN, J. H., and R. R. SOKAL. 1965. A method for deducing branching sequences in phylogeny. *Evolution* **19**:311-326.
- CIFERRI, R., and P. REDAELLI. 1936. Morfologia, biologia e posizione sistematica di *Coccidioides immitis* stiles e delle sue varietà, con notizie sul granuloma coccidioide. *R. Accad. Ital.* **57**:399-474.
- COLE, G. T., and S. H. SUN. 1985. Arthroconidium-spherule-endospore transformation in *Coccidioides immitis*. Pp. 281-333 in P. J. SZANISZLO, and J. L. HARRIS, eds. *Fungal dimorphism with emphasis on fungi pathogenic to humans*. Plenum, New York.
- EDMAN, J. C., J. A. KOVACS, H. MARUS, D. V. SANTI, H. J. ELWOOD, and M. SOGIN. 1988. Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. *Nature* **334**:519-522.

- FARRIS, J. S. 1989. The retention index and the rescaled consistency index. *Cladistics* **5**:417-419.
- FELSENSTEIN, J. 1985. Confidence intervals on phylogenies; an approach using the bootstrap. *Evolution* **39**:783-791.
- GALGIANI, J. N., and N. M. AMPEL. 1990. Coccidioidomycosis in human immunodeficiency virus-infected patients. *J. Infect. Dis.* **162**:1165-1169.
- GALGIANI, J. N., and E. E. WACK. 1988. Coccidioidomycosis. Pp. 1785-1792 in A. P. FISHMAN, ed. *Pulmonary diseases and disorders*, 2d ed. McGraw-Hill, New York.
- GARGAS, A., and J. W. TAYLOR. 1992. Polymerase chain reaction (PCR) primers for amplifying and sequencing nuclear 18S rDNA from lichenized fungi. *Mycologia* **84**:589-592.
- GREUTER, W., H. M. BURDET, W. G. CHALONER, V. DEMOULIN, R. GROLLE, D. L. HAWK-SWORTH, D. H. NICOLSON, P. C. SILVA, F. A. STAFLEU, and E. G. VOSS. 1988. International code of botanical nomenclature. Koeltz Scientific, Königstein.
- HENDRIKS, L., A. GORIS, J. NEEFS, Y. VAN DE PEER, G. HENNEBERT and R. DE WACHTER. 1989. The nucleotide sequence of the small ribosomal subunit of the yeast *Candida albicans* and the evolutionary position of the fungi among the eukaryotes. *Syst. Appl. Microbiol.* **12**:223-229.
- HILLIS, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. Pp. 278-294 in M. M. MIYAMOTO and J. CRACRAFT, eds. *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
- HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.* **83**:189-195.
- JIN, L. I., and J. W. H. FERGUSON. 1990. NJTREE (v2.0). Computer program distributed by the Center for Population Genetics, University of Texas Health Science Center at Houston.
- JORDAN, M. M., J. CHAWLA, M. W. OWENS, and R. B. GEORGE. 1990. Significance of false positive serologic tests for histoplasmosis and blastomycosis in an endemic area. *Am. Rev. Respir. Dis.* **141**:1487-1490.
- KEATH, E. J., E. D. SPITZER, A. A. PAINTER, S. J. TRAVIS, G. S. KOBAYASHI, and G. MEDOFF. 1989. DNA probe for the identification of *Histoplasma capsulatum*. *J. Clin. Microbiol.* **27**:2369-2372.
- KLEIN, B. S., J. M. VERGERONT, R. J. WEEKS, U. M. KUMAR, G. MATHAI, B. VARKEY, I. KAUFMAN, R. W. BRADSHAW, J. F. STOEBIG, J. P. DAVIS, and the Investigation Team. 1986. Isolation of *Blastomyces dermatitidis* in soil associated with a large outbreak of blastomycosis in Wisconsin. *N. Engl. J. Med.* **314**:529-534.
- KURTIN, P. J., D. S. MCKINSEY, M. R. GUPTA, and M. DRIKS. 1990. Histoplasmosis in patients with acquired immunodeficiency syndrome: hematologic and bone marrow manifestations. *Am. J. Clin. Pathol.* **93**:367-372.
- LEE, S. B., and J. W. TAYLOR. 1990. Isolation of DNA from fungal mycelia and single spores. Pp. 282-287 in M. A. INNIS, D. H. GELFAND, J. J. SNINSKY, and T. J. WHITE, eds. *PCR protocols*. Academic Press, San Diego.
- MCGINNIS, M. R., and B. KATZ. 1979. *Ajellomyces* and its synonym *Emonsiella*. *Mycotaxon* **8**:157-164.
- MITCHELL, L. G., and C. R. MERRILL. 1989. Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction. *Anal. Biochem.* **178**:239-242.
- OPHÜLS, M. D. 1905. Further observations on a pathogenic mould formerly described as a protozoan (*Coccidioides immitis*, *Coccidioides pyrogenes*). *J. Exp. Med.* **6**:443-485.
- PAPPAGIANIS, D., and B. L. ZIMMER. 1990. Serology of coccidioidomycosis. *Clin. Microbiol. Rev.* **3**:247-268.
- RIPPON, J. W. 1988. *Medical mycology: the pathogenic fungi and the pathogenic actinomycetes*, 3d ed. W. B. Saunders, Philadelphia.
- RIXFORD, E., and T. C. GILCHRIST. 1896. Two cases of protozoan (coccidioidal) infection of the skin and other organs. *Johns Hopkins Hosp. Rep.* **1**:209-268.

- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SAROSI, G. A., A. CATANZARO, T. M. DANIEL, and S. F. DAVIES. 1988. Clinical usefulness of skin testing in histoplasmosis, coccidioidomycosis and blastomycosis. *Am. Rev. Respir. Dis.* **138**:1082–1083.
- SIGLER, L., and J. W. CARMICHAEL. 1976. Taxonomy of *Malbranchea* and some other hyphomycetes with arthroconidia. *Mycotaxon* **4**:349–488.
- STEVENS, D. A. 1980. *Coccidioidomycosis: a text*. Plenum Medical, New York.
- SUGAR, A. M. 1989. Case 49-1988: *Histoplasma capsulatum* or *Blastomyces dermatitidis*? *N. Engl. J. Med.* **320**:1699.
- SWOFFORD, D. L. 1990. PAUP: Phylogenetic analysis using parsimony, version 3.0. Computer program distributed by the Illinois Natural History Survey, Champaign.
- WHEAT, L. J., P. A. CONNOLLY-STRINGFIELD, R. L. BAKER, M. F. CURFMAN, M. E. EADS, K. S. ISRAEL, S. A. NORRIS, D. H. WEBB, and M. L. ZECHEL. 1990. Disseminated histoplasmosis in the acquired immune deficiency syndrome: clinical findings, diagnosis and treatment, and review of the literature. *Medicine* **69**:361–374.
- WHITE, T. J., T. BRUNS, S. LEE, and J. TAYLOR. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315–322 in M. INNIS, D. H. GELFAND, J. J. SNINSKY, and T. J. WHITE, eds. *PCR protocols*. Academic Press, San Diego.

WESLEY M. BROWN, reviewing editor

Received June 10, 1991; revision received April 21, 1992

Accepted April 22, 1992