Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 519–523, January 1992
Biochemistry

Classification of fungal chitin synthases

(phylogeny/multiple sequence alignment)

A. R. BOWEN*†, J. L. CHEN-WU*‡, M. MOMANY§, R. YOUNG¶, P. J. SZANISZLO§, AND P. W. ROBBINS*‖

*Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; §Department of Microbiology, University of Texas, Austin, TX 78712; and ¶Whitehead Institute for Biomedical Research, Cambridge, MA 02139

Contributed by P. W. Robbins, September 16, 1991

ABSTRACT **Comparison of the chitin synthase genes of *Saccharomyces cerevisiae CHS1* and *CHS2* with the *Candida albicans CHS1* gene (UDP-*N*-acetyl-D-glucosamine:chitin 4-β-*N*-acetylglucosaminyltransferase, EC 2.4.1.16) revealed two small regions of complete amino acid sequence conservation that were used to design PCR primers. Fragments homologous to chitin synthase (≈600 base pairs) were amplified from the genomic DNA of 14 fungal species. These fragments were sequenced, and their deduced amino acid sequences were aligned. With the exception of *S. cerevisiae CHS1*, the sequences fell into three distinct classes, which could represent separate functional groups. Within each class phylogenetic analysis was performed. Although not the major purpose of the investigation, this analysis tends to confirm some relationships consistent with current taxonomic groupings.**

Chitin, the β1-4-linked polymer of *N*-acetylglucosamine, is a fibrous cellulose-like polysaccharide that serves as the major cell wall/exoskeleton scaffolding in many species of fungi, arthropods, insects, and crustacea. In many yeasts chitin is used to maintain the structure of the mother-bud junction, whereas in filamentous fungi chitin is often the major supporting component of the cell wall.

Early enzymatic studies in yeast and filamentous fungi showed that much of the chitin synthase activity was latent, requiring protease activation (1). Recent genetic and molecular studies in *Saccharomyces cerevisiae* suggest that this organism has at least two protease-activated chitin synthase zymogens as well as a more complex chitin synthase system that may not require protease activation (2).

When the derived amino acid sequences of the two *Sa. cerevisiae* chitin synthase zymogens (from *CHS1* and *CHS2*) were compared with the sequence of a closely related *Candida albicans* gene (for UDP-*N*-acetyl-D-glucosamine:chitin 4-β-*N*-acetylglucosaminyltransferase, EC 2.4.1.16), it became clear that the enzymes contained a highly conserved sequence that possibly represents the catalytic region of the enzymes. We decided to use degenerate PCR primers that encoded short, completely conserved sequences within the three genes to probe genomic DNA from a variety of fungi.** PCR-derived fragments were cloned into M13, and single nucleotide sequencing runs were used to classify the clones. Representative clones were then completely sequenced, and the deduced amino acid sequences were put into groups by the CLUSTAL program. The aligned DNA sequences within these groups, or classes, were analyzed further with the FITCH program.

MATERIALS AND METHODS

Fungal genomic DNA was provided by the following laboratories: *Blastomyces dermatitidis* and *Histoplasma capsulatum*, G. Kobayashi (Washington University Medical School); *Aspergillus nidulans*, S. Osmani (Baylor College of Medicine); *Aspergillus niger*, D. Archer (Agricultural and Food Research Council Institute of Food Research, Norwich, U.K.); *Schizophyllum commune*, C. Novotny (University of Vermont College of Medicine); *Ustilago maydis*, J. W. Kronstad (University of British Columbia); *Neurospora crassa*, C. Yanofsky (Stanford University). Gene fragments were cloned into M13 and were grown in DH5αF′ cells.

PCR amplification was performed by using the Perkin–Elmer/Cetus DNA thermal cycler and GeneAmp kit. The primers used were as follows:

```
                    A
                    C
                    G       C   C   A   C
5′ CTG AAG CTT ACT ATG TAT AAT GAG GAT 3′

                  C   A   C   A   A   T
5′ GTT CTC GAG TTT GTA TTC GAA GTT CTG 3′.
```

The conserved sequences encoded by the primers are not included in the analyzed sequences. Thirty to forty cycles were run consisting of a 94°C, 1-min melting step, a 50°C, 1-min annealing step, and a 72°C, 3-min extension. PCR products were digested by *Hin*dIII and *Xho* I, isolated by low-melting-point agarose gel electrophoresis, and inserted into the *Hin*dIII and *Sal* I sites of M13. The chitin synthase gene fragments were sequenced by the method of Sanger *et al.* (3). The reactions were done by using the United States Biochemical Sequenase kit, according to the manufacturer's instructions, with the universal sequencing primer and [α-^{35}S]dATP (New England Nuclear). Several other sequence-specific primers were constructed by the Biopolymer Laboratory at the Massachusetts Institute of Technology. The DNA sequences have been deposited in GenBank.

Computer analysis of the sequence was done by using the programs developed by Roger Staden at the Laboratory of Molecular Biology (Cambridge, U.K.). The amino acid alignment and dendrogram were derived by CLUSTAL (4). The phylogenetic trees were constructed by programs in Joseph Felsenstein's phylogeny inference package PHYLIP 3.2 (5). All programs were run on the VAX/VMS computer through

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: Department of Dermatology, University of Michigan, Ann Arbor, MI 48109.
‡Present address: VICAM, 29 Mystic Avenue, Somerville, MA 02145.
‖To whom reprint requests should be addressed.

**The sequences reported in this paper have been deposited in the GenBank data base [accession nos. M82938-9, *Emericella* (*Aspergillus*) *nidulans*; M82940-1, *Aspergillis niger*; M82942-3, *Blastomyces dermatitidis*; M82944-6, *Exophiala jeanselmei*; M82947-9, *Histoplasma capsulatum*; M82950-1, *Neurospora crassa*; M82952-3, *Phaeococcus exophialae*; M82954-5, *Rhinocladiella atrovirens*; M82956, *Schizophyllum commune*; M82957, *Schizosaccharomyces pombe*; M82958-9, *Ustilago maydis*; M82960-1, *Xylohypha bantiana*; M81905-7, *Wangiella dermatitidis*; M82937, *Candida albicans*].

the computing facility of the Whitaker College of Health Sciences and Technology.

RESULTS AND DISCUSSION

The deduced amino acid sequences for the fragments obtained by PCR amplification are shown in Fig. 1, and a list of their taxonomic affinities and gene designations is presented in Table 1. The amino acid sequences were aligned by the program CLUSTAL (4), which first derives a dendrogram from a matrix of all pairwise sequence similarity scores and then progressively aligns the most similar sequences. The dendrogram (Fig. 2) produced in the first step of the CLUSTAL program was calculated by the unweighted pair group method using arithmetic averages (UPGMA) (10). Although UPGMA dendrograms usually cluster in the appropriate manner, they are not intended to be used as phylogenetic trees (4). Therefore, after the sequence alignment had been obtained, additional methods were consulted to deduce possible phylogenetic relationships.

As pointed out by Valencia *et al.* (11) in their study of ras protein sequences, closeness in sequences such as those considered here "can be interpreted in terms of similarity of function *and/or* in terms of similarity of species." As is clear from inspection of the tree in Fig. 2, both types of similarity are suggested for the chitin synthase gene fragments chosen for analysis. Except for the *Sa. cerevisiae CHS1* fragment, which is left as an "outlying" sequence, the other gene fragments fall into three classes, which could represent three separate functional groups. Although the classification was done by computer, the groups can be found on inspection of Fig. 1 by characteristic gaps and by residues, such as the proline that occurs after the first gap in class III. Within each CLUSTAL class some expected close relationships are seen. For example, similarities are apparent between *A. niger* and *A. nidulans*, which are known or suspected Ascomycetes of the genus *Emericella*, between *H. capsulatum* and *B. dermatitidis*, which are both Ascomycetes of the genus *Ajellomyces*, and among the opportunistic pathogens *E. jeanselmei, W. dermatitidis, P. exophialae*, and *X. bantiana*, which are all members of the same form family (Dematiaciae) of the Fungi Imperfecti, but most likely represent loculomycetous Ascomycetes (7, 9, 12).

To gain a more detailed picture of possible evolutionary relationships, the DNA sequences within each class were compared by programs available in Joseph Felsenstein's phylogeny inference package. The distance matrix programs DNADIST and FITCH were used to produce the class I and class II trees (Figs. 3 and 4). Similarity scores were obtained for all pairwise comparisons of the aligned DNA sequences and were transformed into a distance matrix by the program DNADIST. This program allows for different substitution rates between transitions and transversions, according to the 2-parameter model of Kimura (13). In turn, the distance matrix was used as input to the program FITCH, which calculated branching order and length. *Sa. cerevisiae CHS1*, shown by CLUSTAL to be an outlying species, was used to root the trees.

In addition to confirming the close relationships suggested by the CLUSTAL program, the FITCH tree for class II places the sequences for the two Basidiomycetes *Sp. commune* and *U. maydis* together and separates the single *Ss. pombe* gene fragment into a class by itself (CLUSTAL had grouped *Ss. pombe* and *Sp. commune* together, a result totally inconsistent with modern fungal taxonomic concepts). The most striking aspect of the FITCH analysis, however, is the large evolutionary separation suggested between *Sa. cerevisiae* (and possibly *C. albicans* and *Ss. pombe*) and the other fungi. On morphological grounds, both *Sa. cerevisiae* and *Ss. pombe* are traditionally classified in the same Ascomycete order, Endomycetales (14). Based on a variety of results, *C.*

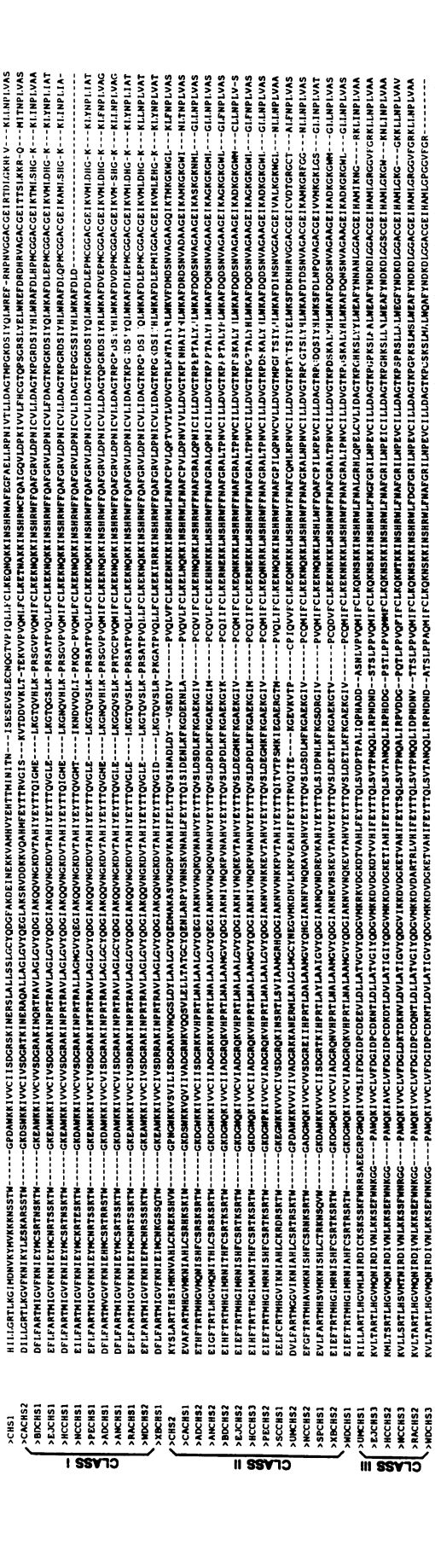

FIG. 1. Multiple amino acid sequence alignment of 32 fungal chitin synthase gene fragments as derived by CLUSTAL. An asterisk indicates complete identity at that position among all fragments, whereas a dot indicates conservative substitutions [defined by Higgins and Sharp (4)]. The NC*CHS1* and WD*CHS3* sequences [first two letters in classes represent the species (see Table 1)] are incomplete as a result of the loss of an *Xho* I and an *Hin*dIII fragment, respectively, during cloning. Only the SC*CHS1* and HC*CHS2* fragments contain introns—3 and 1, respectively. Positions and sequences of the introns are not shown here, but are included in the DNA sequence information deposited in GenBank. The NC*CHS3* sequence (identified in GenBank as chs-1) appears courtesy of Oded Yarden, Stanford University.

Table 1. Taxonomic affinities of fungal species and chitin synthase gene designations

Species	Affinity	Gene designations
Saccharomyces cerevisiae	Ascomycetes	*CHS1, CHS2*
Schizosaccharomyces pombe		Sp*CHS1*
Emericella (Aspergillus) nidulans		Ad*CHS1*, Ad*CHS2*
Aspergillus niger		An*CHS1*, An*CHS2*
Ajellomyces (Blastomyces) dermatitidis		Bd*CHS1*, Bd*CHS2*
Ajellomyces (Histoplasma) capsulatas		Hc*CHS1*, Hc*CHS2*, Hc*CHS3*
Neurospora crassa		Nc*CHS1*, Nc*CHS2*, Nc*CHS3*
Ustilago maydis	Basidiomycetes	Um*CHS1*, Um*CHS2*
Schizophyllum commune		Sc*CHS1*
*Candida albicans**	Fungi Imperfecti	Ca*CHS1*, Ca*CHS2*
Exophiala jeanselmei†		Ej*CHS1*, Ej*CHS2*, Ej*CHS3*
Phaeococcomyces exophialae†		Pe*CHS1*, Pe*CHS2*
Rhinocladiella atrovirens†		Ra*CHS1*, Ra*CHS2*
Wangiella dermatitidis†		Wd*CHS1*, Wd*CHS2*, Wd*CHS3*
Xylohypha bantiana†		Xb*CHS1*, Xb*CHS2*

Taxonomic affinities are modified from Dixon and Fromtling (6).
*Although asexual, this species has many counterparts in the Hemiascomycetidae (7, 8).
†These species are considered to be asexual species of Loculomycetidae or, less often, the Pyrenomycetidae (9).

albicans is also generally thought to be a member of the same order (7, 8). Should *C. albicans*, in fact, be an ascosporic yeast like *Sa. cerevisiae* and *Ss. pombe*, then it might not be surprising that these three fungi exhibit large separations from the remaining fungi investigated. This point, at least as related to *Sa. cerevisiae*, is of interest in the light of the analysis of glyceraldehyde-3-phosphate genes by Smith (15), who postulates enormous evolutionary separation of *Saccharomyces* and related organisms from the filamentous fungi. Numerous other studies with *Sa. cerevisiae* and *Ss. pombe* suggest considerable evolutionary separation between these two hemiascomycetous species, which may be as great as that between *Sa. cerevisiae* and animals (16).

The amplification of a chitin synthase fragment from *Ss. pombe* DNA was unexpected because members of the genus *Schizosaccharomyces* are generally thought to have no chitin in their cell walls. However, a recent paper by Sietsma and Wessels (17) reports the presence of glucosaminoglycan in *Ss. pombe*.

A second aspect of the FITCH analysis, which is quite remarkable is the clustering of fungi, in both the class I and class II analysis, which are traditionally recognized as related. For the class I tree, all the hyphomycetous and melanized Fungi Imperfecti are included in one branch, reflective of their possible loculoascomycetous affinities, whereas the remaining ascomycetous, ascocarpic fungi are included in another branch consisting of only the cleistothecial and the one perithecial species (9, 17). While the exclusive clustering of the two ascocarpic ascomycetous groups is not apparent in the class II analysis, the clustering together of these same fungi with the hyphomycetous, melanized organisms in a single main branch that diverges from a second branch with the two Basidiomycetes and a third branch that encompasses the known or suspected ascosporic yeasts is still compatible with traditional fungal evolutionary schemes (18). The trees presented here also are compatible with studies by Walker (19), who proposes similar groupings of the Ascomycete species based on 5S ribosomal RNA sequences,

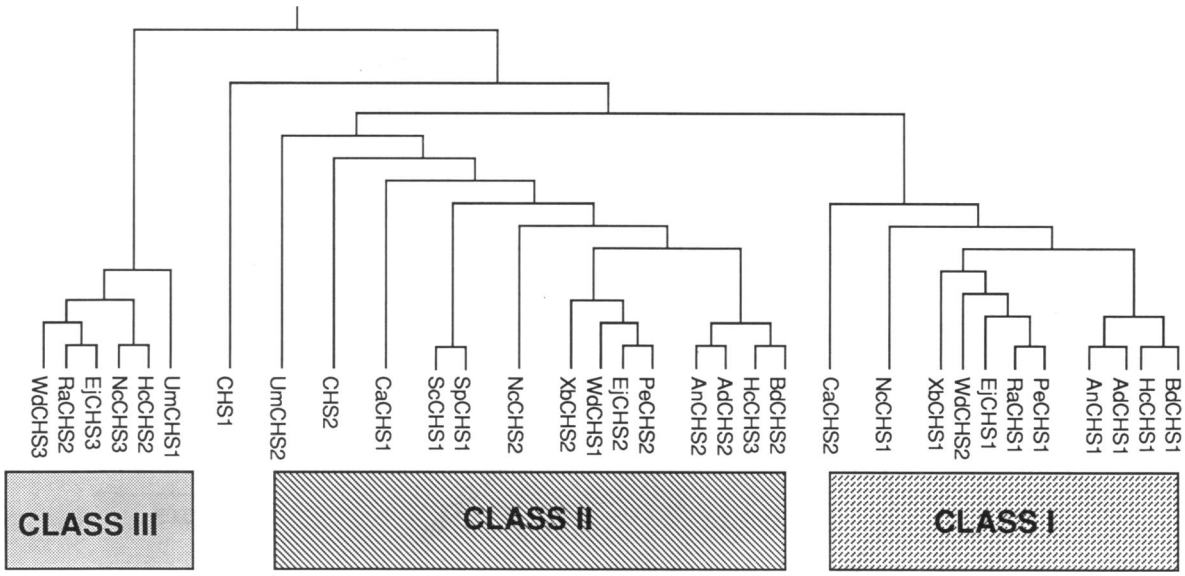

FIG. 2. Unweighted pair group method using arithmetic averages (UPGMA) dendrogram showing three distinct chitin synthase classes. The tree was calculated by the program CLUSTAL from deduced amino acid sequences. Except for *Sa. cerevisiae CHS1*, the gene fragments fall into three groups. Branch lengths do not indicate a rigorous calculation of evolutionary distances, nor can phylogenetic relationships be inferred with confidence from this tree. Except for *CHS1*, the first two letters represent the species (see Table 1).

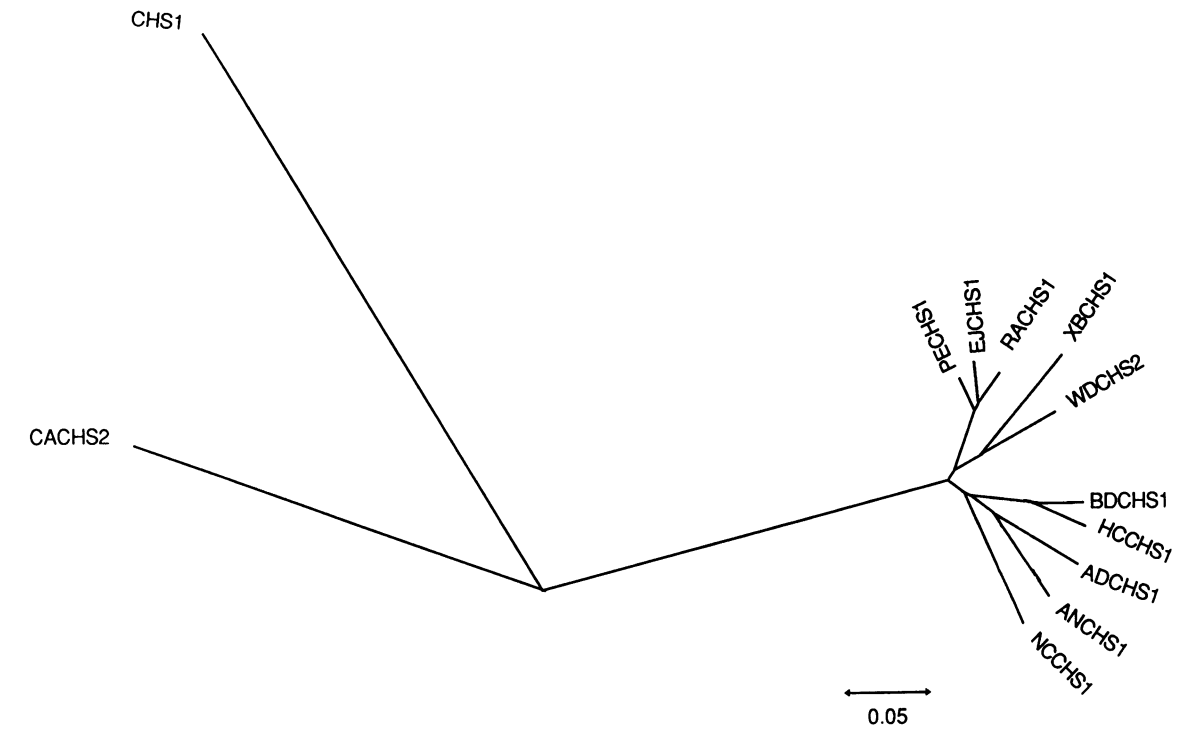

FIG. 3. Tree showing phylogenetic relationships of class I chitin synthase fragments. The DNA sequences were compared by distance matrix methods using the FITCH program. *Sa. cerevisiae CHS1* was included in the analysis as an outgroup to root the tree. Branch lengths reflect relative evolutionary distance and are defined by Felsenstein (5). Except for *CHS1*, the first two letters represent the species (see Table 1).

and with the more extensive tree determined by Hendriks *et al.* (20) by comparison of the small ribosomal subunit RNA sequence. However, it is important to note that the interruption of the Ascomycetes by the Basidiomycete branch in the

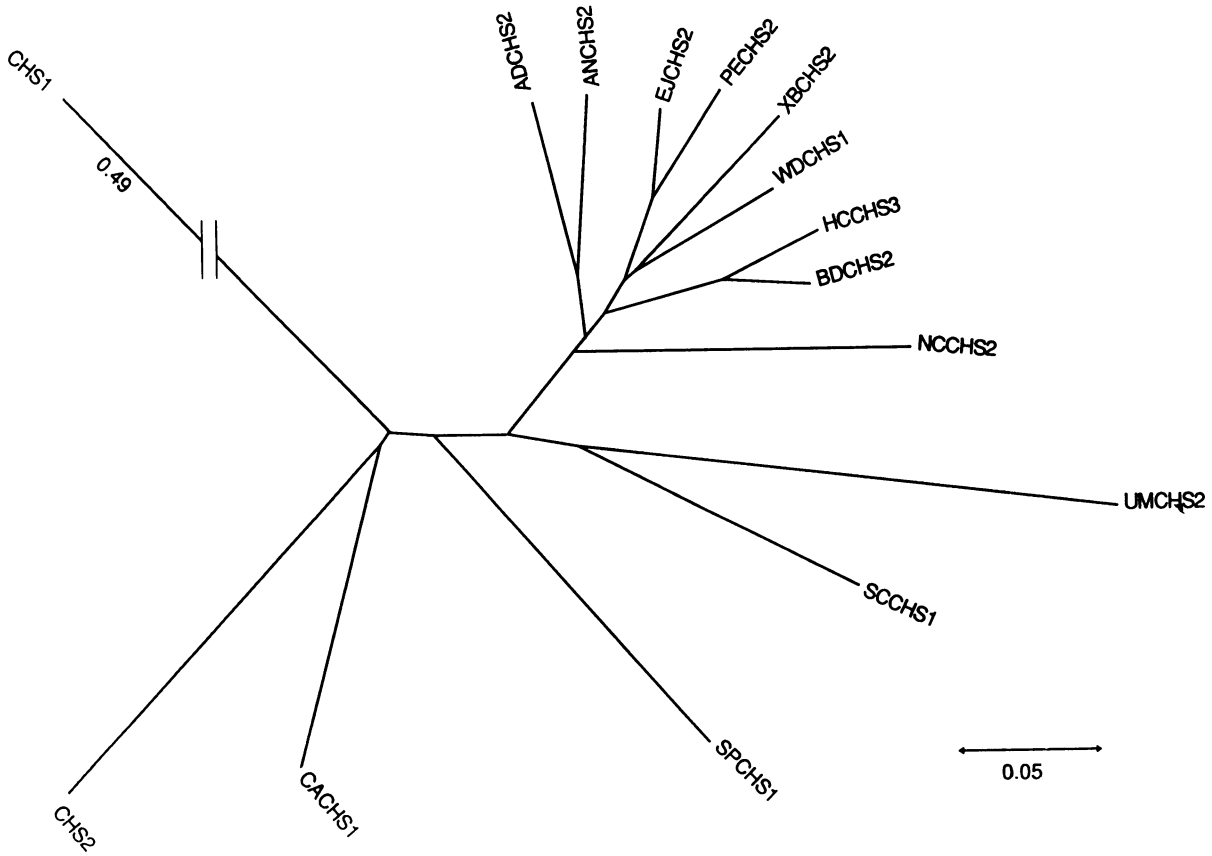

FIG. 4. Tree showing phylogenetic relationships of class II chitin synthase fragments. The DNA sequences were compared by distance matrix methods using the FITCH program. *Sa. cerevisiae CHS1* was included in the analysis as an outgroup to root the tree. Branch lengths reflect relative evolutionary distance and are defined by Felsenstein (5). Except for *CHS1* and *CHS2*, the first two letters represent the species (see Table 1).

class II chitin synthase phylogenetic tree contradicts at least one current sequence analysis of small subunit ribosomal RNAs, which did not suggest that Basidiomycetes are derived from Ascomycetes (21). Possibly this contradiction relates only to the *CHS2* tree being rooted to *CHS1* of *Sa. cerevisiae*. Future analyses using different roots or no roots and additional sequences from other fungi with better established phylogenies should clarify this situation.

Bootstrap analysis was performed by the DNABOOT program (5) as a further means of validating interrelationships inferred by the FITCH analysis. Bootstrapping is a statistical method used to evaluate the confidence level of the phylogenetic estimate by random resampling of the data. One hundred bootstrap replicates were performed on each class, and the resulting consensus confirmed our taxonomic interpretation of the CLUSTAL and FITCH analyses. *E. jeanselmei*, *P. exophialae*, and *R. atrovirens* were grouped together in 95 of the 100 trials, giving them a 95% bootstrap confidence limit as a phylogenetic group. *H. capsulatum* and *B. dermatitidis* appear with a >99% confidence limit as a group. *C. albicans* and *Sa. cerevisiae* appeared as outgroups from the other fungal species in 100% of the bootstrap replicates. Although not showing >95% confidence limits, other close relationships between *U. maydis* and *Sp. commune*, and between the two *Aspergillis* species were predicted by the DNABOOT program, which constructs trees using parsimony methods rather than distance matrix as FITCH does.

Although the sequence analysis reported here suggests the presence of three classes of "zymogen type" chitin synthases in fungi, the results are considered only suggestive and are presented only to serve as a guide for further investigations. Several factors limit the value of our conclusions. In the first place, the analysis was limited to only one very highly conserved region of the genes in question. It will be necessary to extend the analysis to the complete sequences of as many of the genes as possible to determine whether the class and evolutionary relationships suggested here are maintained in the light of complete sequence information. A related point concerns the general function of the protein domain being analyzed. If, as expected, this domain is part of the catalytic region of the enzyme, different classes might differ somewhat in catalytic mechanism, pH optimum, etc. However, this type of variation may or may not be correlated with the biological functions of the enzymes—i.e., catalytic mechanism variants may or may not have different functions in different species.

A second limitation of the data is that all fragments were recovered with a single set of PCR primers. An incomplete set of gene fragments may well bias the analysis toward a subgroup of genes, although we do know that *Sa. cerevisiae CHS1* and *CHS2*, which can be recovered from the genome with our primers, have quite different functions *in vivo* (22). We know that we would almost certainly not recover DNA fragments for the "non-zymogen" chitin synthase III class of enzyme because the *Sa. cerevisiae CSD2* (*CAL1*) gene lacks the sequences used to design our primers (C. Bulawa, personal communication).

A final limitation of our analysis is in the interpretation that can be made of evolutionary relationships. Because, probably for functional reasons, the sequences are all very similar, the apparent evolutionary distances and relationships may be different from those derived by other methods. On the other hand, all chitin synthases may catalyze the same reaction, using the same key residues. The variation seen in the segment examined may, in fact, reflect the accumulation of neutral amino acid changes; therefore, the greater the evolutionary separation, the greater the number of such changes.

In spite of these limitations, we feel that the analysis presented here will be valuable in planning and interpreting gene disruption experiments designed to unravel the functions of the multiple chitin synthase genes in fungi. They should also be useful in conjunction with more extensive sequence data in the analysis of evolutionary relationships among fungi.

We thank the following Massachusetts Institute of Technology undergraduates for PCR fragment isolation and preliminary sequencing results: Monica McConnell, Susan Pauwels, Elly Bulboaca, Banu Ramachandran, Doug Jeffery, Jason Salter, Linda Sun, Wendy Wai, Daniel Wambold, Jork Zwicker, Richard Cheng, Cindy Hummel, Harry Hwang, Rachel McCarthy, and Tracy Kinkaid. This work was supported by Grants GM31318 (to P.W.R.) and CA14051 (to P. Sharp) from the National Institutes of Health. M.M. was supported by grants to P.J.S. from the Texas Applied Technology Program (TATP-4493) and University of Texas Research Institute Program Grant RR07091.

1. Kang, M. S., Elango, N., Mattia, E., Au-Young, J., Robbins, P. W. & Cabib, E. (1984) *J. Biol. Chem.* **259**, 14966–14972.
2. Orlean, P. (1987) *J. Biol. Chem.* **262**, 5732.
3. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
4. Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73**, 237–244.
5. Felsenstein, J. (1989) PHYLIP 3.2 Manual (Univ. of California Herbarium, Berkeley, CA).
6. Dixon, D. M. & Fromtling, R. A. (1991) in *Manual of Clinical Microbiology*, eds. Hawsler, W. J., Jr., Hermann, K. L., Isenberg, H. D. & Shadomy, H. J. (Am. Soc. Microbiol., Washington), 5th Ed.
7. Rippon, J. W. (1988) *Medical Mycology: The Pathogenic Fungi and the Pathogenic Actinomycetes* (Saunders, Philadelphia), 3rd Ed., p. 797.
8. Barns, S. M., Lane, D. J., Sogin, M. L., Bibeaw, C. & Weisburg, W. G. (1991) *J. Bacteriol.* **173**, 2250–2255.
9. deHoog, G. S. & McGinnis, M. R. (1987) in *Proceedings of an International Symposium on the Perspectives of Taxonomy, Ecology and Phylogeny of Yeasts and Yeast-Like Fungi*, eds. deHoog, G. S., Smith, M. T. H. & Weijman, A. C. M. (Elsevier, Amsterdam), pp. 187–199.
10. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy* (Freeman, San Francisco).
11. Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. (1991) *Biochemistry* **30**, 4637–4654.
12. Raper, K. B. & Fennel, D. J. (1973) *The Genus Aspergillus* (Krieger, Huntington, NY).
13. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
14. Lodder, J., ed. (1970) *The Yeasts: A Taxonomic Study* (North-Holland, Amsterdam).
15. Smith, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7063–7066.
16. Nurse, P. (1990) *Cell* **61**, 755–756.
17. Sietsma, J. H. & Wessels, J. G. H. (1990) *J. Gen. Microbiol.* **136**, 2261–2265.
18. Moore-Landecker, E. (1990) *Fundamentals of the Fungi* (Prentice Hall, Engelwood Cliffs, NJ), 3rd Ed.
19. Walker, W. F. (1985) *Syst. Appl. Microbiol.* **6**, 48–53.
20. Hendriks, L., Goris, A., Neeps, J., Van De Peer, Y. & Hennebert, G. (1989) *Syst. Appl. Microbiol.* **12**, 223–229.
21. Illingworth, C. A., Andrews, J. H., Bireal, C. & Sogin, M. L. (1991) *Exp. Mycol.* **15**, 65–75.
22. Bulawa, C. & Osmond, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7424–7428.