Phylogenetic Analysis of *Fusarium* Species Using Ribosomal RNA Sequence Comparisons

Stephen W. Peterson

Microbial Properties Research, National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, Peoria, IL 61604.
Accepted for publication 10 April 1991 (submitted for electronic processing).

The ribosomal RNA genes (rDNA) in fungal cells are organized in a single linkage group as an array of tandemly repeated multigene units. Each of the approximately 100 units present in the haploid genome (Fig. 1) contains the genes for large subunit ribosomal RNA (ls-rRNA), small subunit rRNA (ss-rRNA), and 5.8S rRNA (34) along with transcribed and nontranscribed spacer regions. Sequence homogeneity of rDNA repeat units is maintained for populations of many species (7) apparently through gene conversion (26) and unequal chromosome exchange (reviewed in 7) in the process referred to as molecular drive or concerted evolution. Because of the abundance of rRNA in cells and the conserved sequences contained there, rRNA has been viewed as a prime molecule for phylogenetic reconstruction of organismal histories (35).

In 1985, Lane and associates (23) described a method for rRNA sequencing with reverse transcriptase in a dideoxynucleotide chain termination technique. Crude RNA preparations can be sequenced because the oligonucleotide primers used to initiate the sequence anneal only to particular unique sequences found in ss-rRNA, which is abundant in growing cells. This method is extended to ls-rRNA sequencing by choice of appropriate primers (13). Although the method is rapid, it is limited to single-strand sequencing and some nucleotide positions are occasionally unresolvable. Polymerase chain reaction (PCR) amplification of genes (34) uses similar site-specific oligonucleotide primers and allows sequencing of both DNA strands, extending our ability to rapidly and accurately determine the sequences of particular genes. The advent of these rapid, low-cost techniques for nucleotide sequence determinations of genes has made it possible for many laboratories to use gene sequence comparisons for systematic and taxonomic studies.

Sequence determination of the complete rRNA molecule or its gene (rDNA) is possible, but doing so is time consuming. For purposes of taxonomy and systematics, it may be just as informative to examine and compare only parts of the molecules (23).

Identifying variable sequence regions of rRNA. Secondary structure models (14) and comparisons of nucleotide sequences from diverse organisms show rRNAs to be composed of a core structure along with additional, noncore sequence regions (35). Eukaryotes have sequence regions inserted (16) into the basic rRNA structure that prokaryotes do not have. The sequences in these regions often are less constrained than the core-structure sequences and have been called variable domains and referred to as D regions (17). The patterns of descent among diverse (22,35) or closely related species (12,13) have been determined using sequences from conserved or variable rRNA regions.

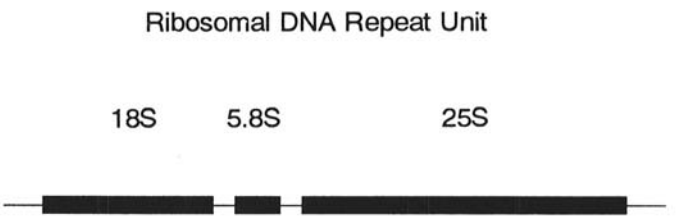

Fig 1. Ribosomal DNA repeat unit, including genes for small subunit (16-18S) rRNA, 5.8S rRNA, and large subunit (23-28S) rRNA that are separated by transcribed and nontranscribed spacer regions (lighter lines).

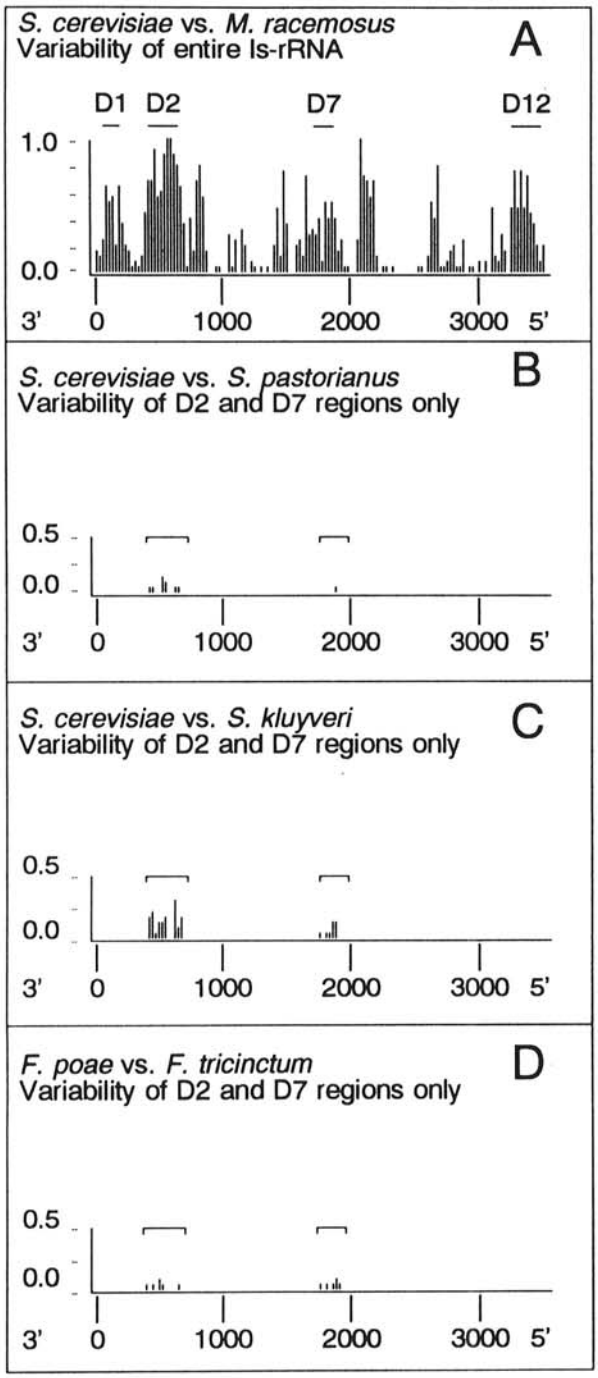

Fig 2. Nucleotide differences between aligned large subunit (ls) rRNA sequences from distantly or closely related fungi. **A,** The proportion of nucleotide mismatches between aligned sequences of *Saccharomyces cerevisiae* (11) and *Mucor racemosus* (19). The D numbers refer to the variable domains defined by Hassouna et al (17) after comparison of ls-rRNA from *Escherichia coli* and five eucaryotic organisms. Alignment in the D2 region is ambiguous due to a 104 nucleotide insert in the *M. racemosus* sequence, but total similarity could not be further improved by alignment changes. **B,** Comparison of D2 and D7 regions of ls-rRNA from the sibling yeast species *S. cerevisiae* and *S. pastorianus* (28) demonstrating greater variability in the D2 than in the D7 region. **C,** Comparison of D2 and D7 regions of the less closely related, but congeneric, species *S. cerevisiae* and *S. kluyveri* (28) showing the relative variability of these regions. **D,** Comparison of D2 and D7 region sequences from *Fusarium poae* and *F. tricinctum* sensu stricto showing the usefulness of D2 region sequences for comparisons of species from *Fusarium* (data from A. Logrieco, S. W. Peterson, and A. Bottalico, *unpublished*).

To examine the variability between two distantly related fungi, the complete ls-rRNA sequences from *Mucor racemosus* (19) and *Saccharomyces cerevisiae* (11) were obtained from GenBank, aligned, and the areas of sequence differences determined (Fig. 2A). *M. racemosus* has an insertion of about 104 bases in the D2 region when compared with the sequence of *S. cerevisiae,* and because the sequences from these two species cannot be unambiguously aligned, meaningful comparison of the species on the basis of D2 region sequences is impossible. However, the great difference between them suggests that the D2 region will be useful for comparisons of more closely related species.

Sequences from four regions of ss-rRNA and two regions of ls-rRNA (200–300 nucleotides per sequence region) were determined and compared to assess intraspecific and interspecific sequence variation (28) among very closely related yeast species. No intraspecific variation was found in the six sequence regions examined. Of two ls-rRNA insertion regions examined (D2, D7), only one would consistently display differences between closely related species (Fig. 2B,C). This region corresponds to the D2 variable domain of ls-rRNA (17) and is about 200 bases long in yeasts and *Fusarium* species. Other regions of ls- and ss-rRNA showed few or no differences between closely related species (28).

Among *Fusarium* species, the D2 region is much more variable than the D1 domain (12) or the D7 domain (Fig. 1D). Sequence variation at the species or strain level also has been found in the spacer regions of rDNA from some organisms (e.g., 25,32), but most sequencing in *Fusarium* has been conducted using rRNA template that does not include the spacer regions. The D2 region of rRNA has been informative for closely related fungi (13,21,24) from several phylogenetically diverse groups. Because the D2 domain sequences are highly variable in many different groups of organisms, it seems a good sequence region to examine when comparing *Fusarium* taxa.

Intraspecific sequence variation. While rRNA sequence homogeneity is maintained in many species (8,12), rDNA from some higher organisms contains sequence heterogeneity in both individuals and species (9,30). Knowing the amount of intraspecific variation is fundamentally important for data interpretation. Because organisms have been found that display different levels of intraspecific sequence variability, it is necessary to determine this characteristic for representative taxa in any group being studied.

Peterson and Logrieco (29) determined partial rRNA sequences (Fig. 3) for seven strains of *Gibberella pulicaris* (Fries) Sacc. (anamorphic state *F. sambucinum* Funkel) and two strains each of intraspecific taxa of *G. fujikuroi* (Sawada) Wollenw. to determine the amount of sequence variation present in the D2 region (17) of species and varieties of *Fusarium. G. pulicaris* has a heterothallic mating system and interfertility has been assessed (4). Vegetative compatibility has been tested among the *G. fujikuroi* varieties studied here (20). The use of fully interfertile strains or those that are vegetatively compatible avoids the difficulties of data interpretation that misidentified strains would cause.

The strains of *G. pulicaris* (*F. sambucinum*), isolated from substrates found on three continents, could be placed in two sequence groups (Fig. 3) that differ from each other at only one nucleotide position. Sequence variability within individual strains was not detected. Some strains with different sequences were isolated from similar geographic regions while some isolates with the same sequence came from different continents. This may reflect the normal variability of this species, but it is also likely that commerce has spread fungal strains found on agricultural products to many areas where natural dispersal might not take them. Those strains with different sequences may represent populations of the fungus that evolved differences under geographic separation without evolving any reproductive isolating mechanisms. Studies of larger numbers of strains from different geographic areas may resolve this question.

Sequence differences (Fig. 3) were not detected among strains of each of the other biological species examined: *G. f. fujikuroi, G. f.* var. *intermedia* Kuhlman, and *G. f.* var. *moniliformis*

(Wineland) Kuhlman. Additionally, no rRNA sequence differences were found between *G. f. intermedia* and *G. f. fujikuroi.* The significance of this will be discussed later.

Guadet and associates (12) examined sequences from the D1 and D2 insertion regions of the ls-rRNA for 51 strains of several taxa with *Fusarium* anamorphs. Several strains of each species were sequenced and 0.35% sequence difference was found between strains of *Nectria haematococca* Berk. & Br. (12), but other species examined had uniform sequence among different strains. Logrieco and co-workers (24) examined the D2 region rRNA sequence of *Michrodochium nivale* (Fries) Samuels & Hallett (formerly *F. nivale*) strains and found no intraspecific sequence variation.

Low intraspecific sequence variation of rRNA sequences suggests that most *Fusarium* species could be identified on the basis of rRNA sequence. Restriction endonuclease fragment length polymorphism (RFLP) studies have detected species-specific, and sometimes strain-specific, bands in fungal biological species (1). Species-specific sequences (seen as rDNA RFLP bands) have been used to differentiate entomopathogenic fungi (15) and sibling species of mosquitos (6). Similarly, oligonucleotide probes that identify particular rRNA sequences are used to identify and quantitate bacterial species in the bovine rumen (31). Techniques for presumptive identification of *Fusarium* strains based on RFLP analysis or oligonucleotide identification of species-specific rDNA sequences may be possible, given the low level of intraspecific variation of rRNA encountered to date.

Phylogenetic analysis. Guadet et al (12) obtained rRNA sequences from 51 strains of *Fusarium* species and some related taxa. Species with *Gibberella* teleomorphs clustered together, and the species with *Nectria* teleomorphs grouped together. *M. nivale* sequences were widely different from any *Fusarium* species sequences. Current morphological distinctions between the species examined by Guadet et al (12) are supported by the lack of sequence variation among strains of these species. Species of teleomorphic genera clustering together adds another confirmation that the generic concepts are valid. *M. nivale,* on the other hand, is so different from either *Nectria* or *Gibberella* species that Guadet et al (12) consider it to be phylogenetically distinct.

The lack of D2 region sequence differences between *G. f. fujikuroi* and *G. f. intermedium* (29) points out limits of resolution in this rRNA sequence analysis. It also suggests that the fertility barriers between these species arose relatively recently and that they are quite closely related phylogenetically. *G. f. moniliforme,* which differs from the varieties mentioned above by two nucleotide substitutions in the D2 region, apparently became genetically isolated before the divergence of the other varieties. This interpretation is supported by the measurements of genetic distance between these varieties made by Ellis using reassociation of nuclear DNA (10).

M. nivale strain NRRL 3289 is difficult to identify morphologically, presumably because it was maintained in culture for many years before its storage in liquid nitrogen (24). It is of interest because it is the only strain of *M. nivale* that is reported to produce trichothecene mycotoxins (referenced in 24). Ribosomal RNA sequences were determined for this strain and for two morphologically typical strains. The sequences of all three isolates were the same. The typical strains have been identified as *M. nivale* var. *major* (24). When these sequences were compared to those of the *M. nivale* var. *nivale* (12) D2 region, there was a sequence difference of about 3%. This information strongly supports the distinction between these taxa. In repeating the mycotoxin analysis, no trichothecene mycotoxins were detected in cultures of strain NRRL 3289 (24).

If rRNA sequences are changing with a clocklike regularity in the regions examined, it should be a good molecule for phylogenetic reconstruction. However, short and long repetitious sequence motifs (16,32,33) have been identified in the D2 variable domain of human rDNA (16) and other higher organisms (9). These are presumably generated by slippage mechanisms (16,33) that would not produce clocklike changes in the sequence. Not all organisms or all insertion regions have these characteristics of slippage-generated sequence mutation and sequence regions should be analyzed for repetitive motifs before use in phylogenetic analysis (9,16,32,33). Kurtzman and Robnett (21) have found different rates of rRNA sequence divergence among genera of yeasts, and thus distance measurements based on D2 region sequences may not be comparable between genera. Introgression of rRNA genes has been detected in insects (2), which adds another

```
383-482                                       #   #                                  #                       #
Reference           .AAAAGAGAG UUAAAAAGUA CGUGAAAUUG UUGAAAGGGA AGCGUUUAUG ACCAGACUUG GGCUUGGUUG AUCAUCUGG* GGUUCUCCCC GGUGCACUCU
G. pulicaris (A)    .......... .......... .......... .........N ..N....... .......... .........A ...N.....* .........N N....N..U.
G. pulicaris (B)    .......... .......... .......... .........N ..N....... .......... .........A ...N.....* .........N A....N..U.
G. f. moniliformis  .......... .......... .......... .........N ..N....... .......... .........A .........* .......... A.......U.
G. f. fujikuroi     .......... .......... .......... .........N ..N....... .......... .........A .........* .........N N.......U.
G. f. intermedium   .......... .......... .......... .........N ..N....... .......... .........A ...N.....* .........N N.......U.
S. cerevisiae       .......... .G........ .......... .......... ..G.CA.U.. .U.....A.. .UG..UUG.. *C.C....CU CC..G.GGGU A.G.G.A...

483-574                          #                                  @                              #
Reference           *UCC*AGCCC *AGGCCAGCA UCAGUUUUCC CCGGGGGAUA AAGGCUUCGG GAAUGUGGCU CUCUUCGGGG AGUGUUAUAG CCCGUUGCGU AAUACC*CUG
G. pulicaris (A)    *...*..U.. *N........ .........G ........N. .......... .......... .C.C...... N......... ...N...U.. ...G..*...
G. pulicaris (B)    *...*..U.. *......... .........G ........NN ...A...... N......... .C.C...... N......... ...N...U.. ...G..*...
G. f. moniliformis  *...*..U.. *......... .........G .......... ...A...... .......... .......... .......... ...N...U.. ......*...
G. f. fujikuroi     *...*..U.. *......... .......... .......... ...A...... .N........ .......... .......... ...N...U.. ......*...
G. f. intermedium   *...*..U.. *......... .......... .......... ...A...... .N........ .......... .......... ...N...U.. ......*...
S. cerevisiae       CG.AUUU.A. UG........ ........GG UG.CA..... ..*U.CAUA. ......A... UG.C....UA ...A...... ..U..*.*.G .....UG.CA

575-624                                                            #      #
Reference           GGGGGGACUG AGGUUCGCG* CAUCUG*CAA GGAUGCUGGC GUAAUGG
G. pulicaris (A)    .C........ .........* .U....*... .......... .......
G. pulicaris (B)    .C........ .........* .U....*.N. .......... .......
G. f. moniliformis  .C........ .........* ......*... ......N... .......
G. f. fujikuroi     .........N .........* ......*... ......N... ..N....
G. f. intermedium   .......... .........* ......*.N. ......N... .......
S. cerevisiae       .CU....... ...AGU...A .G.AA.U... .......... A......
```

Fig. 3. Aligned nucleotide sequences from the D2 region of large subunit rRNA. The reference line is an arbitrary sequence used to facilitate comparison of these data to those of Guadet et al (12). Dots indicate that the species has the same nucleotide as the reference sequence. N indicates the correct nucleotide could not be established. An asterisk is used to indicate spaces added to the sequence to help align the sequences. The # sign is placed over base positions that Guadet et al (12) found readable and contained the same nucleotide for 51 *Fusarium* strains examined, thus suggesting that unresolved bases provide no significant bias in data interpretation. @ marks the *Gibberella pulicaris* heteromorphic site. Sequences were determined by direct sequencing of the rRNA (29) and the species are represented by *G. pulicaris* (A), five strains; *G. pulicaris* (B), two strains; two strains for each of the *G. fujikuroi* varieties; and *Saccharomyces cerevisiae,* six strains. (Data from 28,29.) Numbers above the reference strain refer to the corresponding sequence positions in the published *S. cerevisiae* sequence (11).

problem element to phylogenetic analysis because species of hybrid origin may have patterns of descent that are not indicated by rRNA sequence (18).

Phylogenetic reconstruction of organismal histories using nucleotide sequence data has been performed using distance-matrix calculations (27), parsimony analysis (5), maximum likelihood programs (13), or operator metrics (22). A prime concern, regardless of the analytical method, is the quality of the data, and as Avise (3) points out, multiple sources of data may be needed to confirm phylogenetic histories. Ribosomal RNA sequence data from the D2 region are a good measurement of genetic similarity at the species level and form the basis of phylogenetic hypotheses, but they should be used in conjunction with other data for phylogenetic reconstructions.

The addition of rRNA sequences from more strains and species to our current knowledge of *Fusarium* will provide a fuller framework for determining the genetic affinities of *Fusarium* species. Sequences of rRNA in conjunction with other data, especially from unlinked genes, will provide the basis for forming a phylogenetic reconstruction of the species in this genus.

LITERATURE CITED

1. Anderson, J. B., Petsche, D. M., and Smith, M. L. 1987. Restriction fragment polymorphisms in biological species of *Armillaria mellea.* Mycologia 79:69-76.
2. Arnold, M. L., Shaw, D. D., and Contreras, N. 1987. Ribosomal RNA-encoding DNA introgression across a narrow hybrid zone between two subspecies of grasshopper. Proc. Natl. Acad. Sci. USA 84:3946-3950.
3. Avise, J. C. 1989. Gene trees and organismal histories: A phylogenetic approach to population biology. Evolution 43:1192-1208.
4. Beremand, M. N., and Desjardins, A. E. 1988. Trichothecene biosynthesis in *Gibberella pulicaris:* Inheritance of C-8 hydroxylation. J. Indus. Microbiol. 3:167-174.
5. Bruns, T. D., Fogel, R., White, T. J., and Palmer, J. D. 1989. Accelerated evolution of a false-truffle from a mushroom ancestor. Nature 339:140-142.
6. Collins, F. H., Mendez, M. A., Rasmussen, M. O., Mehaffey, P. C., Besansky, N. J., and Finnerty, V. 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. Am. J. Trop. Med. Hyg. 37:37-41.
7. Dover, G. A. 1982. Molecular drive: A cohesive mode of species evolution. Nature 299:111-117.
8. Dover, G. A. 1987. DNA turnover and the molecular clock. J. Mol. Evol. 26:47-58.
9. Dover, G. A. 1989. Linkage disequilibrium and molecular drive in the rDNA gene family. Genetics 122:249-252.
10. Ellis, J. J. 1988. Section *Liseola* of *Fusarium.* Mycologia 80:255-258.
11. Georgiev, O. I., Nikolaev, N., Hadjiolov, A. A., Skryabin, K. G., Zakharyev, V. M., and Bayev, A. A. 1981. The structure of the yeast ribosomal RNA genes. 4. Complete sequence of the 25S rRNA gene from *Saccharomyces cerevisiae.* Nucleic Acids Res. 9:6953-6958.
12. Guadet, J., Julien, J., Lafay, J. F., and Brygoo, Y. 1989. Phylogeny of some *Fusarium* species determined by large-subunit rRNA sequence comparisons. Mol. Biol. Evol. 6:227-242.
13. Guého, E., Kurtzman, C. P., and Peterson, S. W. 1990. Phylogenetic relationships among species of *Sterigmatomyces* and *Fellomyces* as determined from partial rRNA sequences. Int. J. Syst. Bacteriol. 40:60-65.
14. Gutell, R. R., and Fox, G. E. 1988. Compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res. 16:r175-r270.
15. Hajek, A. E., Humber, R. A., Elkinton, J. S., May, B., Walsh, S. R. A., and Silver, J. C. 1990. Allozyme and restriction fragment length polymorphism analyses confirm *Entomophaga maimaiga* responsible for 1989 epizootics in North American gypsy moth populations. Proc. Natl. Acad. Sci. USA 87:6979-6982.
16. Hancock, J. M., and Dover, G. A. 1991. Compensatory slippage in the evolution of ribosomal RNA genes. Nucleic Acids Res. 18:5949-5955.
17. Hassouna, N., Michot, B., and Bachellerie, J.-P. 1984. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucl. Acids Res. 12:3563-3583.
18. Hillis, D. M., Moritz, C., Porter, C. A., and Baker, R. J. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251:308-310.
19. Ji, G. E., and Orlowski, M. 1990. Primary and secondary structure of the 25s rRNA from the dimorphic fungus *Mucor racemosus.* Curr. Genet. 17:499-506.
20. Kuhlman, E. G. 1982. Varieties of *Gibberella fujikuroi* with anamorphs in *Fusarium* section *Liseola.* Mycologia 74:759-768.
21. Kurtzman, C. P., and Robnett, C. J. 1991. Phylogenetic relationships among species of *Saccharomyces, Schizosaccharomyces, Debaryomyces,* and *Schwanniomyces* determined from partial ribosomal RNA sequences. Yeast 7:61-72.
22. Lake, J. A. 1987. Determining evolutionary distances from highly diverged nucleic acid sequences: Operator metrics. J. Mol. Evol. 26:59-73.
23. Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc. Natl. Acad. Sci. USA 82:6955-6959.
24. Logrieco, A., Vesonder, R. F., Peterson, S. W., and Bottalico, A. 1991. Reexamination of the taxonomic disposition and deoxynivalenol production by *Fusarium nivale* NRRL 3289. Mycologia 83:367-370.
25. Maden, B. E. H., Dent, C. L., Farrell, T. E., Garde, J., McCallum, F. S., and Wakeman, J. A. 1987. Clones of human ribosomal DNA containing the complete 18S rRNA and 28S rRNA genes. Biochem. J. 246:519-527.
26. Nagylaki, T., and Petes, T. D. 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100:315-337.
27. Perasso, R., Baroin, A., Qu, L. H., Bachellerie, J.-P., and Adoutte, A. 1989. Origins of the algae. Nature 339:142-144.
28. Peterson, S. W., and Kurtzman, C. P. 1991. Ribosomal RNA sequence divergence among sibling yeast species. Syst. Appl. Microbiol. 14:124-129.
29. Peterson, S. W., and Logrieco, A. 1991. Ribosomal RNA sequence variation among interfertile strains of some *Gibberella* species. Mycologia 83:397-402.
30. Seperack, P., Slatkin, M., and Arnheim, N. 1988. Linkage disequilibrium in human ribosomal genes: Implications for multigene family evolution. Genetics 119:943-949.
31. Stahl, D. A., Flesher, B., Mansfield, H. R., and Montgomery, L. 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. Appl. Environ. Microbiol. 54:1079-1084.
32. Tautz, D., Tautz, C., Webb, D., and Dover, G. A. 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species. J. Mol. Biol. 195:525-542.
33. Tautz, D., Trick, M., and Dover, G. A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652-656.
34. White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 in: PCR Protocols: A Guide to Methods and Applications. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. Academic Press, New York. 482 pp.
35. Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51:221-271.