Rate and Mode Differences between Nuclear and Mitochondrial Small-Subunit rRNA Genes in Mushrooms[1]

T. D. Bruns and T. M. Szaro
Department of Plant Pathology, University of California, Berkeley

Sequences from homologous regions of the nuclear and mitochondrial small-subunit rRNA genes from 10 members of the mushroom order Boletales were used to construct evolutionary trees and to compare the rates and modes of evolution. Trees constructed independently for each gene by parsimony and tested by bootstrap analysis have identical topologies in all statistically significant branches. Examination of base substitutions revealed that the nuclear gene is biased toward C-T transitions and that the distribution of transversions in the mitochondrial gene is strongly effected by an A-T bias. When only homologous regions of the two genes were compared, base substitutions per nucleotide were roughly 16-fold greater in the mitochondrial gene. The difference in the frequency of length mutations was at least as great but was impossible to estimate accurately because of their absence in the nuclear gene. Maximum likelihood was used to show that base-substitution rates vary dramatically among the branches. A significant part of the rate inconstancy was caused by an accelerated nuclear rate in one branch and a retarded mitochondrial rate in a different branch. A second part of the rate variability involved a consistent inconstancy: short branches exhibit ratios of mitochondrial to nuclear divergences of <1, while longer branches had ratios of ~4:1–8:1. This pattern suggests a systematic error in the branch length calculation. The error may be related to the simplicity of the divergence estimates, which assumes that all base positions have an equal probability of change.

Introduction

The application of direct RNA sequencing and the recent development of PCR primers for rRNA genes and spacers has made sequence analysis of these regions increasingly popular in fungal studies (Guadet et al. 1989; Barnes et al. 1991; Bruns et al. 1991). The technology for acquiring these sequences, however, is much more advanced than our understanding of the relative rates and modes of evolution within these regions. Specifically, the base-substitution-rate difference between mitochondrial and nuclear genes has not been examined in fungi, nor has the frequency of different types of base substitutions. Rate and frequency data could aid in the selection of appropriate regions for phylogenetic studies and could also be incorporated into decisions on character-state weighting.

The purpose of this paper is to compare sequence evolution in portions of the mitochondrial and nuclear small-subunit ribosomal RNA (SrRNA) genes. The regions chosen overlap homologous parts of domain 2. Prior comparative studies have shown

1. Key words: transition-transversion ratio, likelihood, base composition effects, molecular clock, strand bias.

Address for correspondence and reprints: T. D. Bruns, Department of Plant Pathology, University of California, Berkeley, California 94720.

Mol. Biol. Evol. 9(5):836–855. 1992.

0737-4038/92/0905-0007$02.00

that parts of these regions are among the most variable in the SrRNA gene (Gutell et al. 1985; Sogin and Gunderson 1987). In distant phylogenetic comparisons these portions are typically unalignable (Gray et al. 1984; Cedergren et al. 1988). Other parts of the regions studied are highly conserved and have been used to infer the origin of mitochondria and the phylogeny of Eukaryotes (Gray et al. 1984; Cedergren et al. 1988).

The 10 fungi we investigated are members of the Boletales, an order of mushroom-producing basidiomycetes. We chose this group because our prior work suggested that they span a low to moderate range of molecular diversity (Bruns et al. 1989; Bruns and Palmer 1989). This represents a relatively uninvestigated level of divergence in fungi, but it is one that we expected to be useful for studying the evolution of these genes.

We first used parsimony analysis to analyze the two genes independently and as a combined data set. We then assessed confidence in these trees by comparing the topology of the nuclear and mitochondrial trees and by bootstrapping. The best parsimony-analysis tree was then used to examine the frequency of different types of base substitutions. These frequencies were then used as an aid in the selection of appropriate parameters for maximum-likelihood analyses. The likelihood analyses were used to examine the constancy of nucleotide substitution rates.

Material and Methods

DNAs from *Boletus satanas* (TDB-1000), *Gomphidius glutinosus* (TDB-957), *Chroogomphus vinicolor* (TDB-1010), and *Paxillus atrotomentosus* (TDB-782) were extracted from freeze-dried collections according to a method described elsewhere (Bruns et al. 1990). The voucher collections will be deposited in the University of California herbarium. DNA extraction methods and deposition of voucher collections for other fungal taxa—*Suillus sinuspaulianus* (DAOM-66995), *S. cavipes* (TDB-646), *Paragyrodon sphaerosporus* (TBD-420), *Phylloporus rhodoxanthus* (TDB-540), *Rhizopogon subcaerulescens* (F-2882), and *Xerocomus chrysenteron* (TDB-635)—have been given elsewhere (Bruns et al. 1990; Bruns and Palmer 1989).

All sequences except those from the mitochondrial SrRNA of *S. sinuspaulianus* were obtained by direct sequencing of PCR products by the asymmetric-primer-ratio method (Gyllensten and Erlich 1988), with minor modifications described elsewhere (Bruns et al. 1990). The NS3/NS4 primers at ratios of 50:1 and 1:50 and the ML1/ML2 primers at ratios of 50:2 and 2:50 were used to generate templates for both strands of the nuclear and mitochondrial genes, respectively (White et al. 1990). One additional primer—MS4, AACCACCATTCATCGTTGAC—was used to sequence the mitochondrial gene, and two additional primers—NS20UCB, CGTCCCTATTAATCATTACG; and NS21UCB, GAATAATAGAATAGGACG—were used to sequence the nuclear gene. The latter two primers were designed by Gargas and Taylor (1992). Sequencing was done with *Taq* polymerase according to the manufacturer's (USB) instructions. Sequences from the mitochondrial SrRNA of *S. sinuspaulianus* were obtained from cloned regions of native DNA. Clones and sequencing methods have been described elsewhere (Bruns et al. 1989; Bruns and Palmer 1989).

Initial sequence alignments were done with Genalign (Intelligenetics) by using equal weights for deletions and mismatch, a length factor of 6, and a spread factor of 50. Final alignments for the mitochondrial sequences were visually adjusted.

The nuclear and mitochondrial data were analyzed independently and in combination by uniformly weighted parsimony analysis using the branch-and-bounds algorithm of PAUP (Swofford 1990), with gaps treated as missing data. Confidence intervals on tree topologies were estimated by bootstrap analyses (Felsenstein 1985)

by using the heuristic search option of PAUP. The nuclear data were used unaltered and in their entirety, but, within the mitochondrial sequences, not all regions were used, because of difficulties with alignment (figs. 1 and 2). The mitochondrial V8 region was completely omitted from the analysis. The mitochondrial V7 region was used by double coding the sequences as follows: the aligned block of the first five taxa (fig. 2) was used with positions for the other five taxa encoded as missing data, and then the aligned block for the last four taxa (fig. 2) was entered, with the first six taxa's positions coded as missing. The *Paxillus* V7 region, which could not be aligned with any of the other taxa, was coded as missing in both blocks. The small ambiguous region in V7 of *Paragyrodon* was also coded as missing in both blocks. This approach enables five additional cladistically informative sites, as well as 16 additional unique changes, to be included in the analysis. Inclusion of these sites affects topology among the most closely related species and also contributes to terminal branch lengths within all taxa except *Paxillus*. The only disadvantage in this approach is that the lengths of some longer internal branches were underestimated relative to most terminal branches, because the long internal branches connect taxa that could not be aligned within the V7 region. When rates were compared (see below), this problem was corrected by omitting the V7 region.

Results

Distribution of Sequence Differences

Aligned sequences of the 10 taxa are shown for the nuclear and mitochondrial regions studied (figs. 1 and 3). Variation within the nuclear sequence was almost entirely limited to the eukaryote-specific (ESR) and V8 regions. The ESR region is an area of unresolved structure found only in eukaryotic nuclear SrRNAs, and the V8 region is a helical region known to be variable, in shape and size, over broad phylogenetic comparisons (Gutell et al. 1985). Only five differences were observed outside these two regions, and none was found in the "universal" regions sampled. No length mutations of any size were observed in the nuclear genes.

In contrast, variable sites were distributed throughout the mitochondrial genes, including portions of the universal regions. Although the universal regions are noticeably more conserved than the surrounding sequences, 19% of the sites in these supposedly conserved regions are variable within this order of fungi. In addition, many length mutations were observed in the mitochondrial gene. Most were located in the two highly variable regions discussed below, but at least seven small length differences were found in other regions (fig. 1).

The two most variable areas of these mitochondrial genes, V7 and V8 (figs. 1 and 2), correspond to the 588–651 and 829–857 regions of the *Escherichia coli* SrRNA gene. The first region is absent in eukaryotic nuclear SrRNA genes but typically forms stem structures in bacterial and most mitochondrial genes; these stems are known to be variable, in shape and size, among phylogenetically distant taxa (Gutell et al. 1985). In the Boletales the V7 region can potentially be folded into helical structures, but the structural variation within the order is virtually as great as that in all previously described fungal mitochondrial SrRNAs (fig. 4). Many apparently uncompensated changes are tolerated, but most are localized in two bulge-loop regions (fig. 4). Neither the V7 region nor the V8 region of the mitochondrial gene could be unambiguously aligned by primary structure in all 10 taxa; V7 is alignable within two subsets of the taxa, while only fragmented patterns of similar sequences exist in V8 (fig. 2). Both regions are strongly biased toward adenine and thymine, averaging 86% and 92%

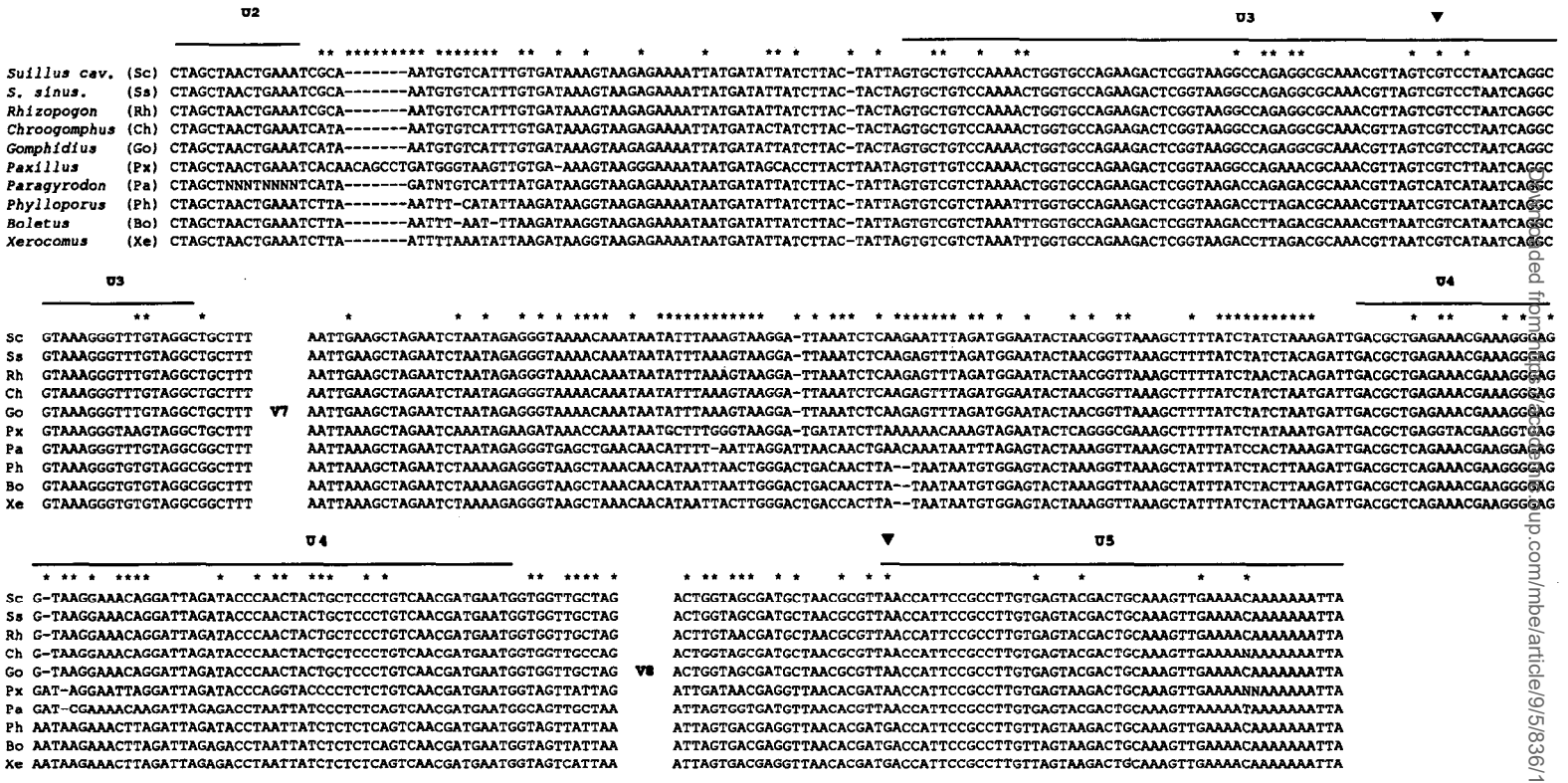

FIG. 1.—Sequence alignments from a portion of the mitochondrial SrRNA gene. The sequences start 14 bp from the MS1 primer and end 32 bp before the MS2 primer. Variable positions are indicated with an asterisk (*). The positions of the "Universal" regions U3, U4, and U5 and of variable regions V7 and V8, as defined by Gray et al. (1984) and extended by Cummings et al. (1989), are indicated, but the sequences for the V7 and V8 are omitted from this alignment and are given in fig. 4. The black, downward-pointing triangles (▼) indicate the borders of the area homologous to the aligned nuclear sequences (fig. 3). A dash (–) indicates a gap introduced for alignment. Full species names and isolate numbers are given in Material and Methods. GenBank accession numbers for the sequences (from top to bottom) are M91016, M91017, M91015, M91010, M91011, M91012, M91014, M91013, M91009, and M91018.

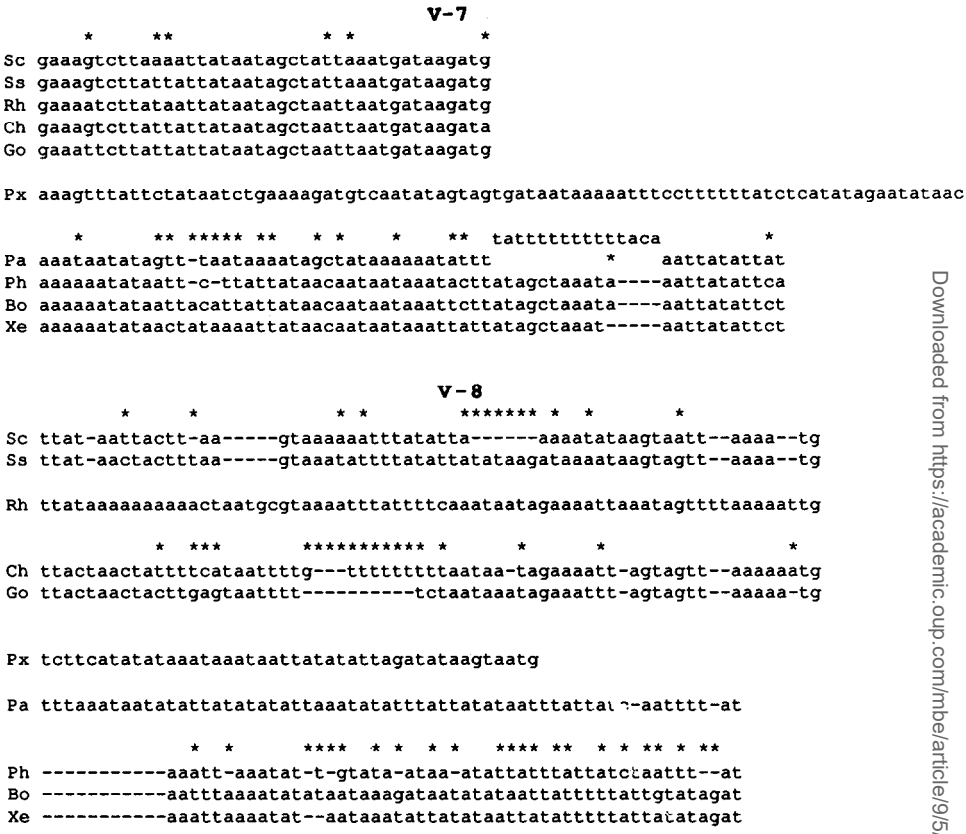

FIG. 2.—Partial alignments of the hypervariable regions (V-7 and V-8) from the mitochondrial SrRNA gene. Rows of adjacent sequences are shown aligned for each of the two variable regions. Within the V-8 region, nonadjacent rows are only roughly aligned among the suilloid taxa (Sc, Ss, Rh, Ch, and Go) and between the boletoid taxa (Ph, Bo, and Xe) and *Paragyrodon* (Pa); other nonadjacent blocks are not aligned. An unaligned portion of the V-7 region of *Paragyrodon* is shown out of register. Variable positions within aligned rows are indicated by asterisks (*).

A+T, compared with ~60% for most other areas of the genes. Together these two regions account for the overwhelming majority of the length differences among these sequences.

Phylogenetic Analyses

Trees derived independently from the nuclear and mitochondrial data were identical in all statistically significant portions of their topologies (fig. 5), and both demonstrate, with considerable confidence, that the family Boletaceae is polyphyletic. In both trees, one member of the family, *Paragyrodon,* is the sister taxon to three species that we will refer to as the "boletoid group," but only in the mitochondrial data is this relationship demonstrated with statistical confidence. *Suillus,* the only other representative of the Boletaceae, was found to be most closely related to *Rhizopogon* (Hymenogastraceae), *Gomphidius,* and *Chroogomphus* (Gomphidiaceae); collectively we will refer to these five genera as the "suilloid group." *Paxillus* was found to be a distant sister taxon to the suilloid group, in both trees (fig. 5).

Neither the mitochondrial nor the nuclear data resolve, with significant levels of

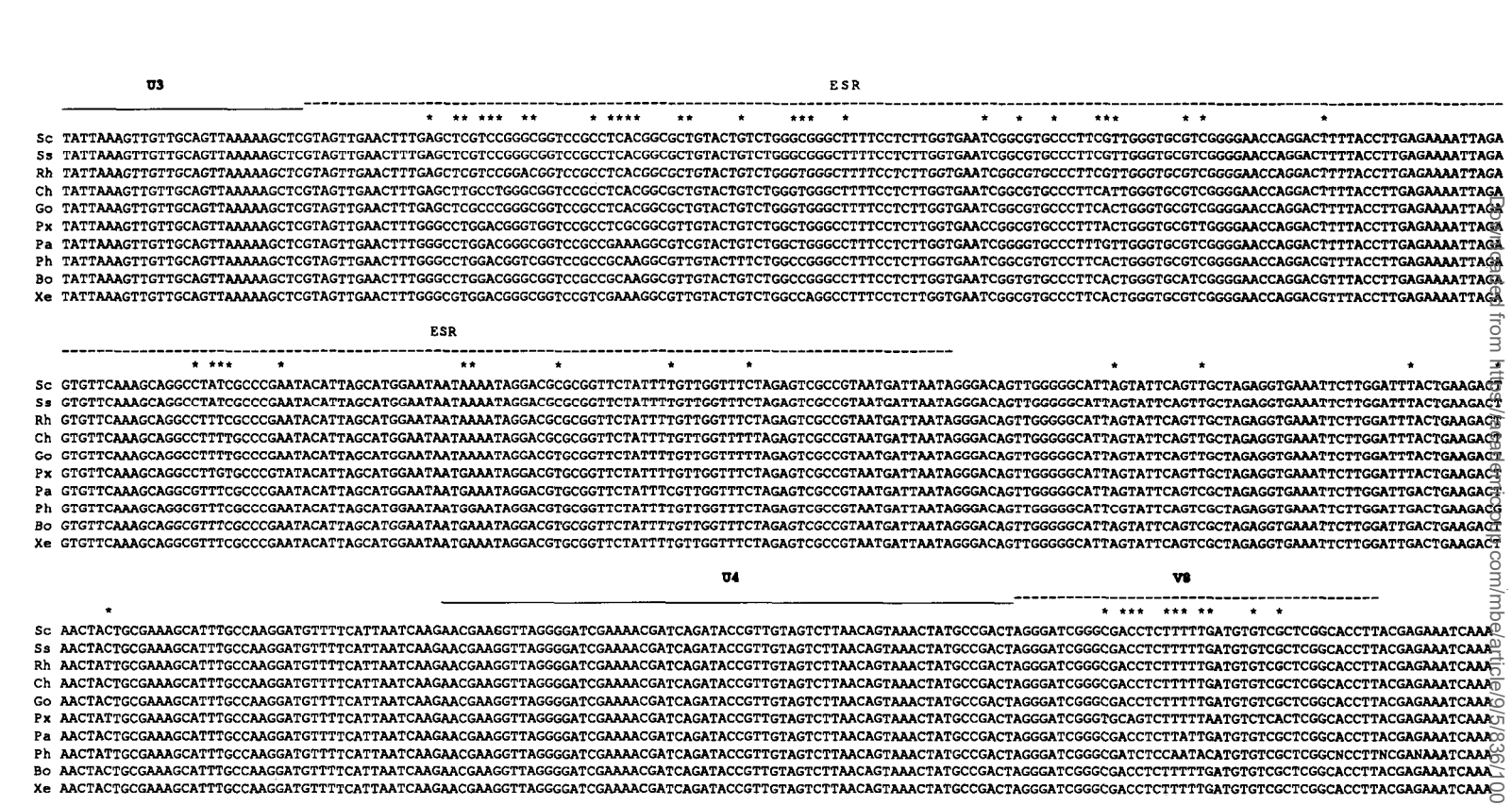

FIG. 3.—Sequence alignments from a portion of the nuclear SrRNA gene. The sequences start 16 bp from the NS3 primer and end 40 bp before the NS4 primer. The U5 region starts immediately after the sequence shown. A eukaryotic-specific region (ESR) (Gutell et al. 1985) is indicated. Symbols and abbreviations are as in fig. 1. GenBank locus numbers (from top to bottom) are M90828, M90829, M90827, M90822, M90823, M90824, M90826, M90825, M90821, and M90830.

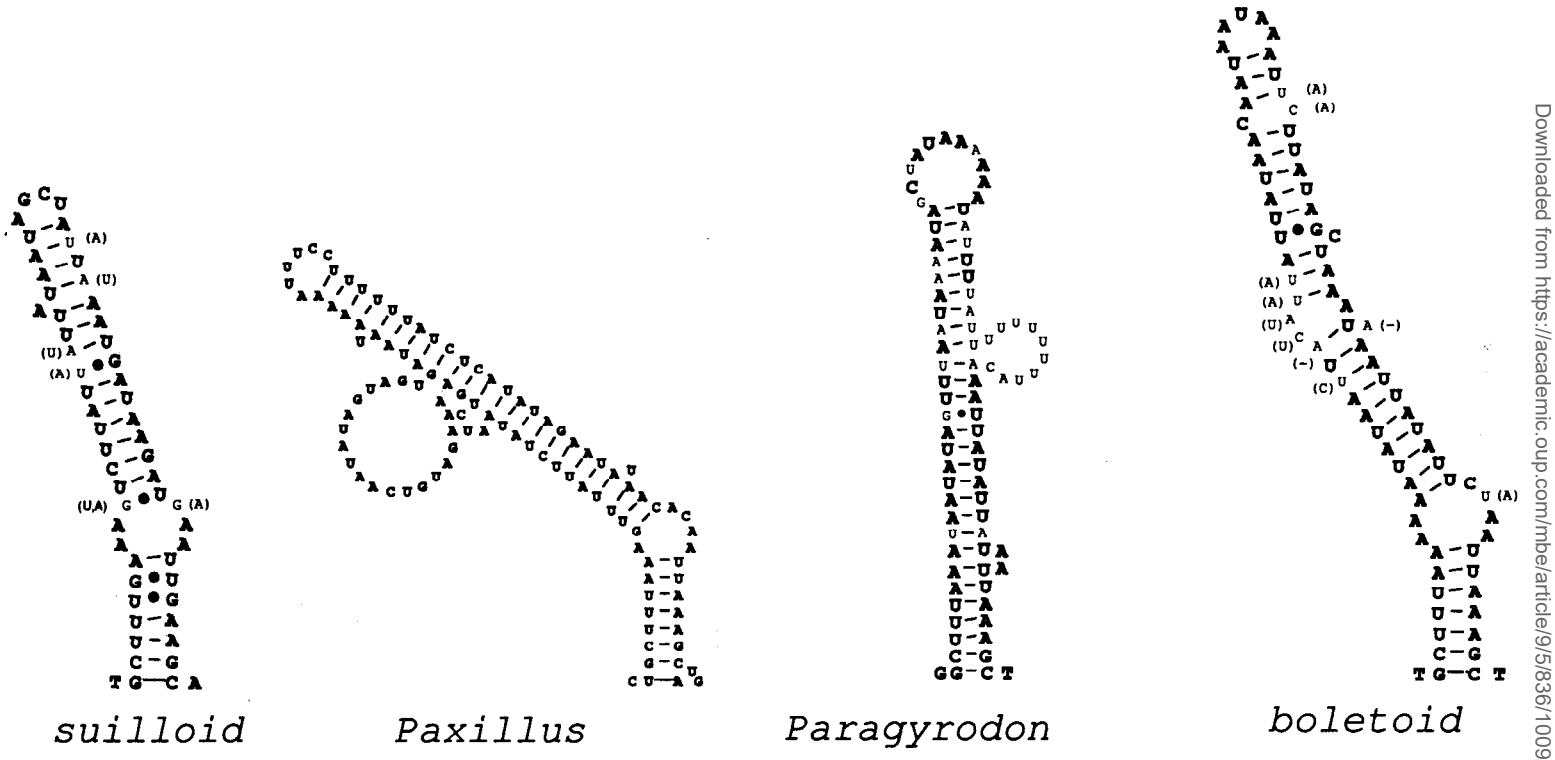

FIG. 4.—Possible stem structures formed by the V7 region and flanking sequences. The suilloid and boletoid structures show the sequences of *Suillus sinuspaulianus* and *Boletus satanas,* with bases that differ among members of their respective groups shown in lightface type. Base substitutions in the other taxa within the boletoid and suilloid groups are shown in parentheses. Lightface type in the *Paragyrodon* structure indicates bases that differ from those in the boletoid group. Structures were identified with PCfold2 (computer program written by Michael Zucker and Daniel Brunelle, National Research Council of Canada, Ottawa) by constraining the flanking conserved regions into the structures found in other small-subunit genes. Other structures with similar predicted stabilities also exist.

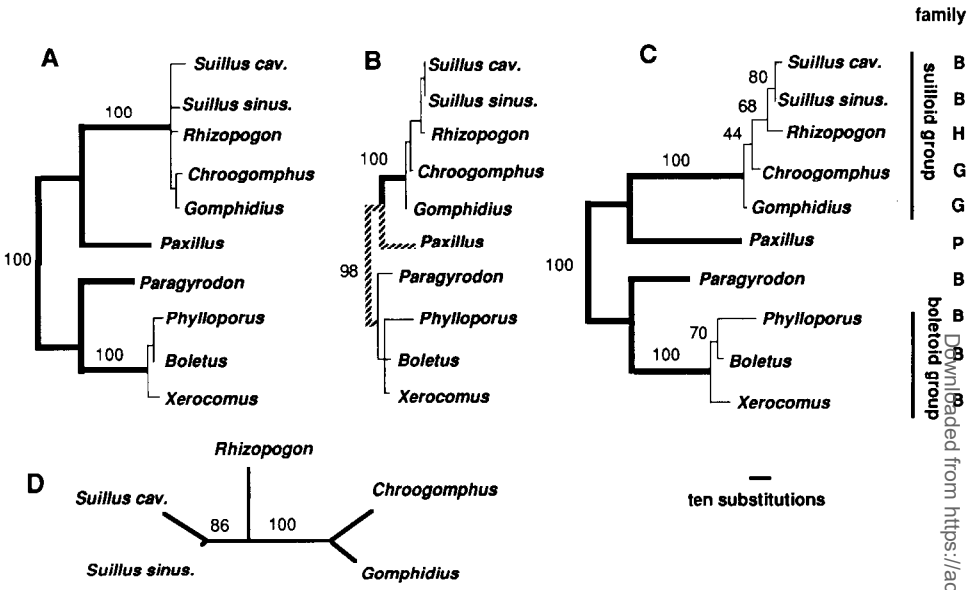

FIG. 5.—Phylogenetic analysis. Trees bases on mitochondrial (A), nuclear (B), and combined (C) data sets are drawn to the scale shown. Handling of ambiguous alignments in the mitochondrial gene are discussed in the text. Tree A is the consensus of two equally parsimonious trees. Trees B and C are the shortest trees found. Topologies depicted in boldface type were present in >99.9% (*thicker solid black lines*) or >98% (*striped lines*) of the trees found in a bootstrap sample of 10,000 trees; numbers indicate the percentages of trees in the bootstrap sample that support the indicated branches. Percentages are not shown for branches in trees A and B that were supported by <95% of the trees. The boletoid and suilloid groups are indicated along with traditional family classifications: B = Boletaceae; G = Gomphidiaceae; H = Hymenogastraceae; and P = Paxillaceae. Tree D is drawn to a fivefold greater scale; it shows results of a bootstrap analysis (10,000 replicates) of the unrooted suilloid group, by use of the combined data. This tree is also the most parsimonious for either data set alone. PAUP 3.0L was used for all trees.

confidence, the relationships within the boletoid or the suilloid groups. In both cases the ambiguity is not really centered on topology within these groups but, instead, on how these topologies are rooted by the rest of the tree. For example, within the boletoid group there is only one possible three-taxon tree, but the mitochondrial and nuclear data root it differently relative to *Paragyrodon,* and in neither data set is the topology specified with confidence. Similarly, within the suilloid group, both the mitochondrial and nuclear data result in a single most parsimonious tree for the five taxa (fig. 5D), but the rooting of this suilloid tree relative to *Paxillus* differs. When the mitochondrial and the nuclear data were pooled, statistical support for the internal branch separating the two members of the Gomphidiaceae from the other suilloids can be demonstrated (fig. 5D), but which taxa are monophyletic remains unresolved (fig. 5C).

Base-Substitution Biases

By taking tree C (fig. 5) as our best estimate of the phylogeny, we observed the frequency of base substitutions by examining the character-state changes inferred by the parsiomony analysis (fig. 6). This approach is justified even if tree C is not correct, because errors are very likely limited to the shortest branches, as the four longest branches are strongly supported by bootstrap analysis and are also congruent between the separate gene trees. Rearrangements of topology in the short branches result in

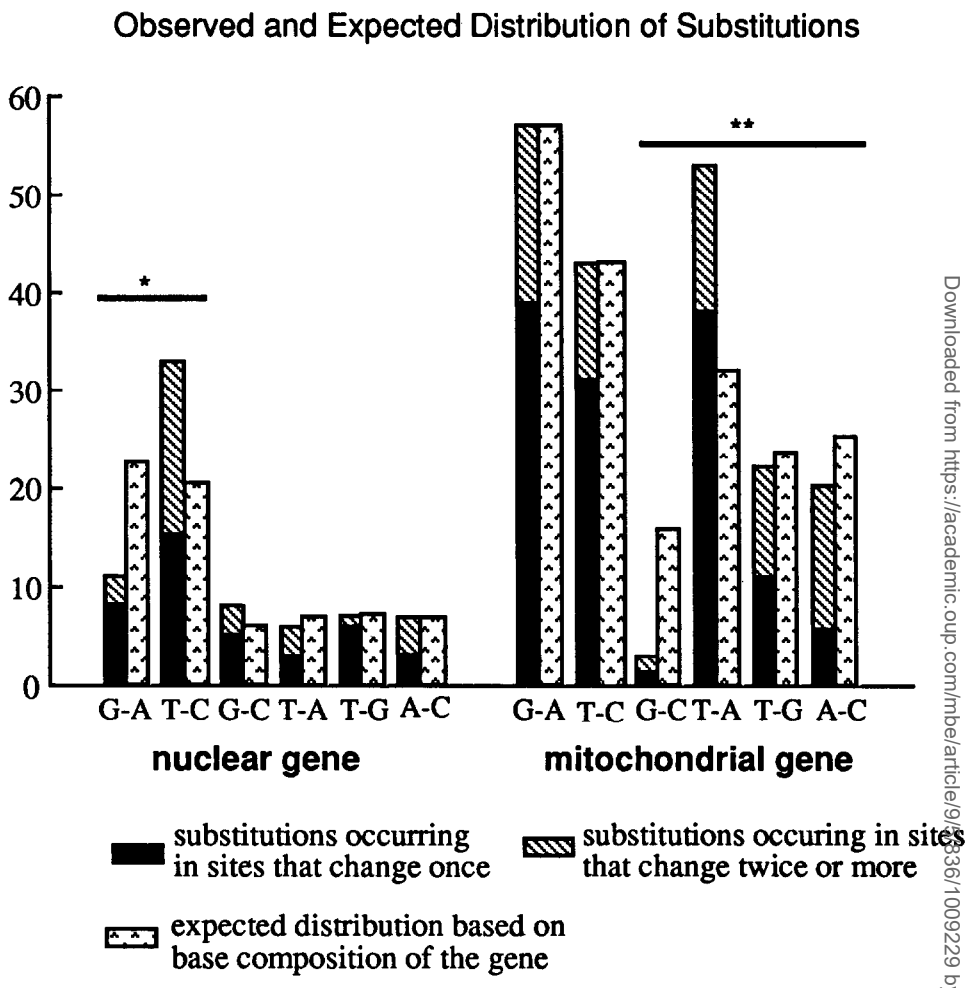

FIG. 6.—Observed and expected distributions of substitutions. Observed substitutions are inferred from character-state changes on tree C (fig. 5). Expected substitutions were calculated separately for transition and transversion by assuming that the probability of observing a particular type of substitution is proportional to the frequency of bases involved. Observed values that differ significantly from expected are indicated (χ^2 * = $P < 0.05$; and ** = $P < 0.01$).

very minor differences in the number of specific base substitutions. The directionality of substitutions would be affected (e.g., C to T vs. T to C), but, because the tree as a whole is unrooted, we did not attempt to infer directionality.

Both the nuclear gene and the mitochondrial gene exhibit a slight bias toward transitions, with the nuclear gene appearing slightly more biased (fig. 6 and table 1). It has been shown elsewhere, however, that strong transition biases may only be observable when sequence divergence is very low. For example, in *Drosophila* mtDNA, a strong bias was observed only among sequences with <2% divergence (DeSalle et al. 1987). To look for evidence of such a strong cryptic bias, we examined substitutions among the least divergent portions of the tree (i.e., the suilloid and boletoid groups), and we partitioned the changes between sites that were observed to change only once and those that were observed to change two or more times. We also determined a

Table 1
Observed and Estimated T Values

	T[a] (no. of transitions/no. of transversions)	
	Nuclear Gene	Mitochondrial Gene
Observed:		
Entire tree	1.7 (45/27)	1.0 (100/98)
Suilloid group only	10 (10/1)	1.4 (7/5)
Boletoid group only	1 (10/10)	0.4 (2/5)
Suilloid and boletoid combined	1.8 (20/11)	0.9 (9/10)
Among sites observed to change only once	1.3 (22/17)	1.2 (69/56)
Maximum-likelihood estimate [95% confidence interval]	1.5 [0.9–2.4]	1.3 [1–1.7]

[a] Observed ratios are derived from examining inferred character-state changes in tree C in fig. 2. Maximum-likelihood estimates were obtained from the top tree in fig. 7A, as described in the text. 95% confidence intervals were determined from likelihood-ratio tests.

maximum-likelihood estimate and 95% confidence interval for the transition/transversion ratio (T) for a subset of six taxa (see below). The only indication that a stronger cryptic bias may exist is the 10:1 ratio observed in the nuclear SrRNA sequences within the suilloid group. This particular comparison, however, is the only one in which the taxa have ~2% sequence divergence; all others are ≥3%–4% different. This latter level can be sufficient to obscure a highly biased T value if the value is determined by pairwise comparisons or counting changes inferred from parsimony analysis (DeSalle et al. 1987). The 10:1 ratio in the suilloid group is significantly different (χ^2; $P < .05$ level) from the treewide observed ratio of 1.7 but is not significantly different from ratios ≥1.8. This latter value is well within the likelihood estimate of T, which should be robust to branch length distortions.

Among the transitions in the nuclear gene, C-T substitutions predominate. The observed ratio of G-A transitions to C-T transitions differs significantly (χ^2; $P < 0.05$) from that expected from base-compositional biases of the gene (fig. 6). This deviation necessitates a strand-specific bias, because C-T transitions that occur on the coding strand will appear as G-A transitions on the noncoding strand.

Our observed C-T bias might be explained in at least three ways: (1) a high rate of C-T changes could result from deamination of 5-methylcytosine; (2) C-T changes in helical regions of the gene may often be selectively neutral, because both C and U can pair with G, and thus preserve stems; and (3) by chance, a high percentage of the least-constrained sites are occupied by C or T in this region of the nuclear gene. Explanation (1) is attractive because deamination of methylcytosine is the most frequent single-base mutation in eukaryotic cells and is implicated in the fungal process known as RIP, which results in an extremely high mutation rate in duplicated regions (Cambareri et al. 1989). It would also explain why the C-T bias is not seen in the unmethylated mitochondrial gene. In order for deamination of methylcytosine to account for the observed predominance of C-T transitions on the noncoding strand, methylation would have to be either more common on the antisense than on the sense strand or, at least, more prone to deamination. However, we examined the distribution of potential methylation sites (i.e., CpG dinucleotides) and found the numbers of sites to be identical on both strands in the nuclear sequences. If only explanation (2) is

correct, it does not explain why the C-T bias is not seen in the mitochondrial gene. Both explanation (2) and explanation (3) explain why a high percentage of C-T changes occur at sites that have changed more than once (fig. 5). Both predict that other nuclear regions should not exhibit this bias, at least if they are also not constrained by the secondary structure of their RNA products [explanation (2) only]. To test this prediction, we examined the partial 28S sequences from *Fusarium* (Guadet et al. 1989), the glyceraldhyde-3-phosphate dehydrogenase genes of *Saccharomyces* (Smith 1989), and internal transcribed spacer sequences of *Suillus grevillei* (Baura et al. 1992). All three regions contained a strong C-T transition bias on the noncoding strand. The bias in these latter examples suggests that a more general mechanism for strand bias exists in the nuclear genomes of fungi. Berbee and Taylor (accepted) have now confirmed that, within the 18S gene of ascomycetes, the constraints of secondary structure are not likely to be the cause of this bias. Initial examination, by Gojobori et al. (1982), of the substitution patterns in vertebrate pseudogenes revealed just the reverse bias: G-A transitions were much more frequent on the noncoding strand. But in a later, expanded study this bias disappeared (Li et al. 1984). In any case, both of these later studies suggest that the observed C-T bias may be limited to fungi.

The observed ratios of transversions in the mitochondrial gene also differ significantly (χ^2; $P < 0.01$) from values predicted from the base composition of the gene (fig. 6), but this departure from expected values is caused exclusively by the high frequency of T-A substitutions and the low frequency of G-C substitutions. The deviation becomes much less pronounced if a higher A+T percentage is assumed. At 85% the deviation is no longer significant at the 0.01 level, and at 90% it is no longer significant at the 0.05 level. The most variable regions of the gene (i.e., V7 and V8) have base compositions in this range, suggesting that such skewed compositions may provide a more appropriate estimate of the random nucleotide replacement pool in the mitochondrial gene.

Rate Differences

The rate of base substitution is clearly much greater in the mitochondrial gene, as evidenced by the branch lengths of the nuclear and mitochondrial trees in figure 5. This comparison is not entirely appropriate, however, because some regions used from each gene have no homologue in the other. When nonhomologous and unalignable regions are excluded (the nonoverlapping regions of the mitochondrial gene, the ESR region from the nuclear gene, and the V7 and V8 from both genes), then the total lengths of the nuclear and mitochondrial trees are 7 and 129, respectively. After correction for the slight difference in number of nucleotides compared, the rate difference translates to 16-fold. This value probably underestimates the true difference because parsimony analysis makes no multiple-hit correction. This omission would be expected to result in a greater underestimate of the longest (i.e., mitochondrial) branches.

Length mutations were completely ignored in the above rate estimate, but their inclusion would further increase the estimated rate difference. No length differences were observed in any portion of the nuclear gene, while in the mitochondrial gene two regions were rendered unalignable by the frequency of such events (fig. 2). One of these regions, the V8 stem, does have a homologue in the nuclear gene. Poor alignment makes accurate estimation of the number of length mutations nearly impossible in the V8 region, but it is probably in excess of 15 events (fig. 2). If we restrict the

comparison to the same homologous regions used for estimating base-substitution rates, then the difference in frequency of length mutations is less pronounced (3:0).

Rate Constancy

If the base-substitution rate were relatively constant in both genes, one would also expect the base-substitution ratio between the two to be constant. However, a visual comparison of the parsimony-analysis trees (fig. 5A and B) suggests that significant variation has occurred in the relative substitution rates of the mitochondrial and the nuclear genes: the longest branches show a large rate difference, while the short branches are very similar in length, in both the mitochondrial and nuclear trees. The branch leading to *Phylloporus* is actually longer in the nuclear tree. To examine whether this apparent rate variation is real, we used maximum likelihood to estimate branch lengths. The advantage of using maximum likelihood is that it is easily adapted to the estimation of branch lengths on a known tree, and the parameters used in the likelihood model can be altered to assess how robust the conclusions are to assumptions about T and base composition (Felsenstein 1981). Identical trees produced by different assumptions can also be statistically compared by a likelihood-ratio test (Felsenstein 1990).

The tree topology used is shown in figure 7A. Note that several branches within the suilloid and boletoid groups were pruned in order to eliminate the problem of the undetermined topologies within these two clades and to reduce the computational time for the likelihood analyses. The topology of this tree was strongly supported by bootstrap analysis of the mitochondrial gene and by congruence between the parsimony analyses of the two genes. All regions alignable in both the mitochondrial gene and the nuclear gene were used, since the question we were addressing was not the absolute rate difference but, rather, the constancy of rates between two fixed (though not strictly homologous) regions. For this subset of taxa, the T values that maximized the likelihoods were determined by iteration, and the 95% confidence intervals around these maxima were determined by a likelihood-ratio test (Felsenstein 1990). T values determined in this way were used in the initial likelihood analyses. The results show that no single rate line can intersect all points (fig. 7B). Branches 7 and 8 are the only statistically significant outliers, but the best estimates for the four shortest mitochondrial branches (i.e., 1, 2, 8, and 9) show a mitochondrial/nuclear base-substitution ratio of <1.0, compared with the other branches, rate ratios, which vary from ~ 2 to 10. This result is not unique to the likelihood divergence estimates. Pairwise comparisons among the same six taxa by means of Kimura's (1980) two-parameter model also produce rate ratios that vary from 0.2 to 6.1, and again the shortest branches are the only ones with ratios <1. Parsimony analysis also yields similar results.

The likelihood branch length estimates are of course dependent on both the value of T and the base compositions used. Because the accuracy of our T estimates is uncertain for both data sets, and because the appropriate values for base composition are unclear in the mitochondrial data, we examined the effect of both parameters on branch length. T was found to have no significant effect on the nuclear estimates, but higher values of T did cause slightly elevated estimates of the longer internal mitochondrial branch lengths (fig. 7C). Thus the rate variation may actually be underestimated if the mitochondrial value of T is greater than the value chosen in our analysis. High A+T values were found to dramatically decrease estimates of the longer mitochondrial branches (fig. 7C), but the overall pattern of rate inconstancy remains the same (fig. 7B, *bottom*).

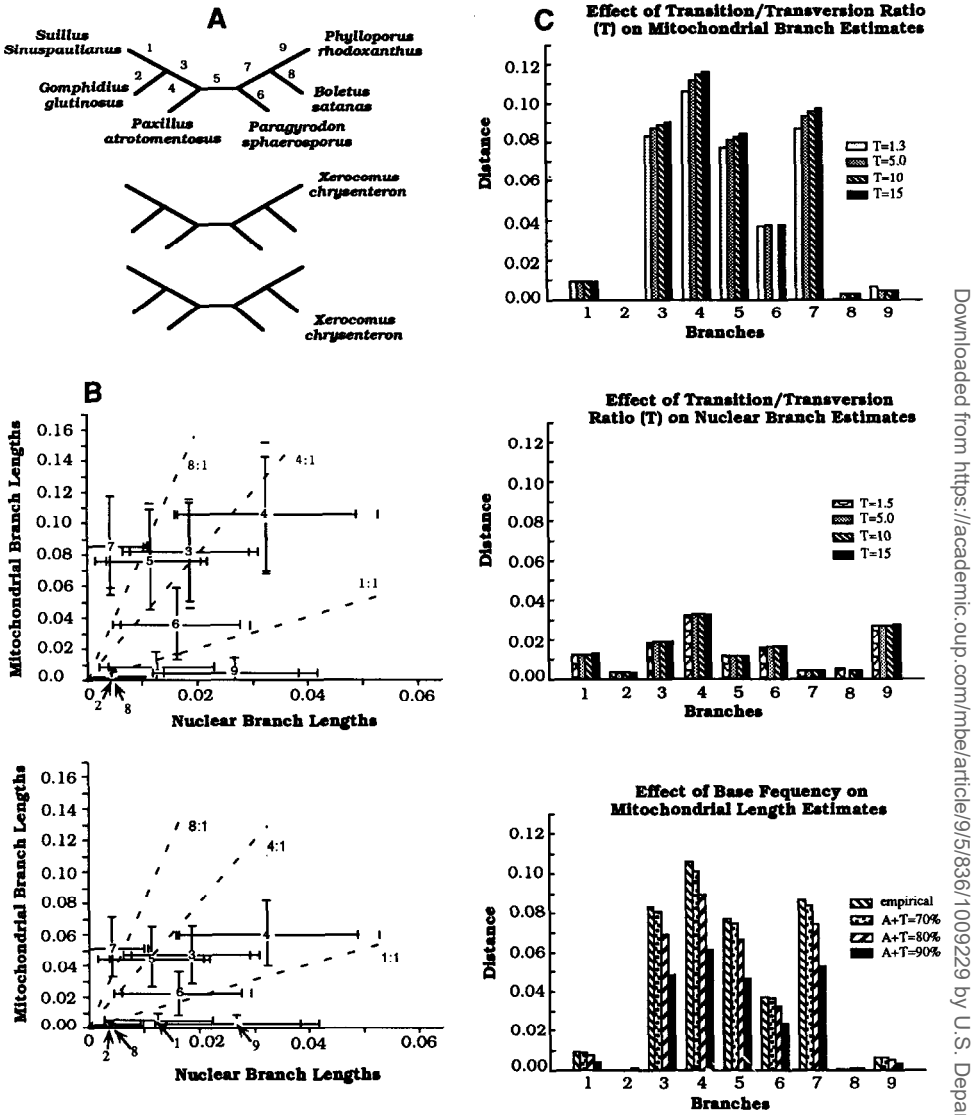

FIG. 7.—Maximum-likelihood analyses. A, Three trees. The top tree was used for most maximum likelihood analyses; its numbered branches are used in the remainder of the figure. The lower two trees were used only in tests of the molecular clock (see table 2) and differ by the substitution of *Xerocomus* for the two other members of the boletoid group. B, Maximum-likelihood estimates of mitochondrial expected branch lengths and nuclear expected branch lengths, plotted against each other. The dotted lines show the ideal rates indicated. The top graph uses empirical base frequencies, while the bottom graph uses a 90% A+T value for the mitochondrial estimates. Error bars indicate 95% confidence intervals. Both the rough calculated error estimates (terminal attached lines) and a bootstrapped error estimate (unattached or non-terminal lines) are shown. The latter was based on a 100-tree sample generated by SEQBOOT version 3.4 (J. Felsenstein, personal communication). Only the rough calculated error estimates were obtained for the 90% A+T values. C, Effects that parameters used have on branch length estimates.

The variation in relative rate implies that one or both genes are not evolving in a clocklike manner. To examine this possibility, we used a likelihood-ratio test to compare the results of likelihood analysis with and without a molecular-clock constraint (Felsenstein 1990). In the clock model the tree topology was not constrained, since the root of the tree was unknown, but all trees produced by the clock model had a topology identical to that used in the nonclock model (see fig. 7A, *top*). Several comparisons were made for both trees, using different values of T. For the mitochondrial gene different A+T contents were also tested. We also combined the data, to see whether local inconstancies of one gene would be offset by those of the other. Rejection of the likelihood-clock model was possible for the initial tree (fig. 7A, *top*) under all conditions tested, except when an A+T content of 90% was assumed for the mitochondrial gene (table 2). When the two genes were analyzed independently, all clock-based trees produced a root along the branch leading to *Paxillus*. The combined data also specified the same root when the mitochondrial site and the nuclear site were treated as a single class (i.e., a mitochondrial/nuclear base-substitution ratio of 1). When the combined data were analyzed with a two-rate-class model having a mitochondrial/nuclear base substitution ratio of 4:1, the root shifted slightly, to branch 5

Table 2
Tests of Likelihood Molecular-Clock Model under Various Conditions

Test Parameters[a]	Likelihood Value[b]		Rejection of Clock[c]
	Without Clock	With Clock	
Mitochondrial tree T:			
T = 1.3; empirical base frequency	−1,283.34	−1,291.03	Yes ($P < 0.01$)
T = 1.3; 70% A + T	−1,294.56	−1,301.84	Yes ($P < 0.01$)
T = 1.3; 80% A + T	−1,331.45	−1,338.61	Yes ($P < 0.01$)
T = 1.3; 90% A + T	−1,446.20	−1,449.77	No ($0.25 > P > 0.1$)
T = 10; empirical base frequency	−1,347.93	−1,354.93	Yes ($P < 0.01$)
Mitochondrial tree M:			
T = 1.3; empirical base frequency	−1,292.76	−1,303.51	Yes ($P < 0.001$)
Nuclear tree T:			
T = 1.5; empirical base frequency	−1,060.44	−1,065.60	Yes ($P < 0.05$)
T = 10; empirical base frequency	−1,082.73	−1,087.98	Yes ($P < 0.05$)
Nuclear tree M:			
T = 10; empirical base frequency	−1,012.30	−1,013.72	No ($P > 0.5$)
Nuclear tree B:			
T = 10; empirical base frequency	−1,072.63	−1,076.80	No ($0.1 > P > 0.05$)
Combined, tree T:			
T = 1.5; empirical base frequency			
Single-rate-class model (mitochondrial/nuclear = 1)	−2,410.91	−2,422.91	Yes ($P < 0.001$)
Two-rate-class model (mitochondrial/nuclear = 4)[d]	−2,410.91	−2,572.90	Yes ($P < 0.001$)

[a] Tree letters T, M, and B refer to top, middle, and bottom trees, respectively, in fig. 7A.
[b] Calculated with DNAML and DNAMLK programs from versions 3.3 of PHYLIP (Felsenstein 1990).
[c] Likelihood-ratio test with 4 degrees of freedom, as described in the PHYLIP manual (Felsenstein 1990).
[d] Ratio chosen from fig. 7B.

near the node leading to *Paxillus.* Other roots were not tested but would result in a lower likelihood and, therefore, in a stronger rejection of the clock model.

Our initial rate comparison (fig. 7B) identified the branches leading to the boletoid group (i.e., branch 7) and *Phylloporus* (i.e., branch 9) as the only significant outliers. To examine whether the source of this variation was mitochondrial or nuclear, we compared the internal rate constancy of these two branches by a relative-rate test. If rates are constant within the trees, then branch 7 should equal the difference of branches 6 and 9 or of branches 6 and 8, and branch 9 should equal branch 8. When these comparisons were made using the likelihood estimates (fig. 7B), branch 7 was found to be ~2.4–2.8-fold longer than expected, in the mitochondrial tree, and branch 9 was found to be 6.4-fold longer than expected, in the nuclear tree. It is relatively simple to test whether the rate acceleration seen in branch 9 of the nuclear tree is the primary reason for the rejection of the clock. This can be done by substituting an excluded member of the boletoid group (i.e., *Xerocomus*) for *Phylloporus,* in the likelihood analyses. When this was done with the nuclear tree, the results showed that the clock model and the nonclock model yielded very similar likelihood values and, thus, that rejection of the clock is no longer possible (table 2). In the mitochondrial tree the substitution of this taxon does not change the rejection of the likelihood clock, thus demonstrating that the effect is caused by an acceleration of the nuclear substitution rate in *Phylloporus.* Unfortunately, the effect of the nuclear rate acceleration in *Phylloporus* is not independent of which boletoid member it is paired with. When *Xerocomus* is substituted for *Boletus* (table 2), the significance of the clock deviation drops below the 95% level, although not very far below it.

In spite of both the rejection of the likelihood-clock model under a variety of parameters and the sensitivity of branch length estimates to both base composition and T, we found that all of the maximum-likelihood trees produced with the clock model were topologically identical to the subset tree produced from parsimony analysis. This result reinforces Fukami-Kobayashi and Tanteno's (1991) study that showed that the topology of likelihood trees is relatively robust to different input parameters, even though branch length estimation is not.

Discussion

The higher rate of mitochondrial base substitution seen in these fungi is reminiscent of that observed in animals (Brown et al. 1982) but contrasts with the slow rate found in plants (Wolfe et al. 1987). However, the question of whether the difference in substitution rates in these highly constrained genes parallels differences in the neutral mutation rates remains to be determined. In any case, the fast mitochondrial rRNA rate has at least two practical implications. First, for those interested in phylogenetic analysis of fungi, the mitochondrial rRNA genes may provide a better choice than nuclear rRNA genes, for addressing questions at intermediate levels of divergence. Second, for those addressing questions of very deep divergence, the mitochondrial gene may well be out of range. The relatively deep levels of divergence found within the "universal" regions of the gene suggest that many of the variable positions within these regions are likely to be saturated by multiple substitutions when more distantly related organisms are compared. The inherent noise resulting from saturation may well explain how the same data can be differently interpreted with respect to the question of mitochondrial origins (Gray et al. 1984; Van de Peer et al. 1990).

What is the basis of the higher mitochondrial substitution rate in fungi? Differences in both mutation and fixation rates can clearly contribute. The latter could be a function

of reduced constraints, owing to the more limited function of the mitochondrial ribosome and its removal from the selection of the "error catastrophe loop"—negative feedback of translation errors into DNA replication and repair (Wilson et al. 1985). The observation that even the so-called universal regions of the mitochondrial gene appear to change rapidly fits well with a model of reduced constraints, although faster mutation rates cannot be eliminated as a factor. Differences in the mode of transmission could also have an impact. The apparently uniparentally inherited mitochondrial gene would be more prone to both reduction in diversity and acceleration in rate of molecular divergence, via bottlenecks (DeSalle and Templeton 1988). Even though recombination can occur in fungal mitochondrial genomes, concerted evolution might be rather different from that in the nuclear rDNA, because the multicopy mitochondrial genomes are easily compartmentalized into separate mitochondria and are segregated mitotically.

Rate variability seems to have at least two components in these genes. The most significant component involves large rate differences in specific branches: branches 7 and 9 in the mitochondrial and the nuclear tree, respectively. In the case of the nuclear gene we have shown that the rate acceleration in branch 9 is the primary cause for rejection of the likelihood-clock model. Other lineage-specific rate variations, of similar magnitude, have been noted elsewhere (Li et al. 1987). The reasons for these variations in branches 7 and 9 remain unknown, but, whatever their causes, the two genes clearly responded to them differently.

Rejection of the likelihood clock and rejection of the molecular clock are not necessarily synonymous. The likelihood-clock model requires that the expected branch lengths be equal among all lineages, from the root to the tips of the tree. This model is rejected if the nonclock model yields a likelihood that is 2 standard deviations greater. The likelihood-ratio test of the clock also assumes equal rates of evolution at all sites. Variation in nucleotide sequences is known to be greater than would be expected from a Poisson distribution (Langley and Fitch 1974). Differing interpretations of this observation may accommodate constant rates (Gillespie and Langley 1979). As a result, adherents of the molecular clock have long accepted a "looser" clock (Wilson et al. 1977). The relative-rates variation, between the nuclear gene and the mitochondrial gene, that is seen in our data (fig. 7B) shows that, even if one is not willing to reject the molecular clock, times derived from divergence estimates of these sequences may be virtually worthless.

The second part of the variability involves a consistent inconstancy: shorter branches appear to have greater nuclear divergence, while longer branches appear to have much greater mitochondrial divergence. This pattern suggests that at least this part of rate inconstancy is artifactual. The work of Palumbi (1989) suggests a possible source of the artifact, one that is consistent with the idea of reduced constraint in the mitochondrial genes. Palumbi showed that, even if the rate of change is constant in two genomes, the ratio of divergence between them can vary directionally over time if the percentage of positions free to vary differs between the genomes compared. His example is based on animals, where the faster mutation rate in the mtDNA initially results in a higher rate of divergence for that genome. Because there are fewer positions free to vary in mtDNA, nuclear divergence becomes greater than mtDNA divergence, when more distantly related organisms are compared. To apply Palumbi's explanation directly to these fungal SrRNA genes, we would have to postulate that the rate of sequence divergence in the nuclear gene is actually greater than that in the mitochondrial gene—but that the number of positions free to vary is much lower in the nuclear

gene. The faster nuclear rate, which is necessary for this interpretation, could be an artifact resulting from the particular portions of the two genes used. Recall that in these comparisons the most variable regions were omitted from the mitochondrial gene (i.e., V7 and V8) but were included in the nuclear gene (i.e., ESR and V8).

A problem with the above explanation is that the likelihood analysis should have corrected for saturation. One reason that it may not have done so is that different classes of sites would be expected to saturate at different times (fig. 8), but the model we used treated all sites as a single class. In view of the well-recognized differences in conservation of various regions of rRNA (Gray et al. 1984; Gutell et al. 1985; Sogin and Gunderson 1987), a single-rate-class model is clearly unrealistic. We have avoided a multiple-rate-class model, however, because of the difficulty associated with defining both the classes and the relative probabilities of substitution between classes.

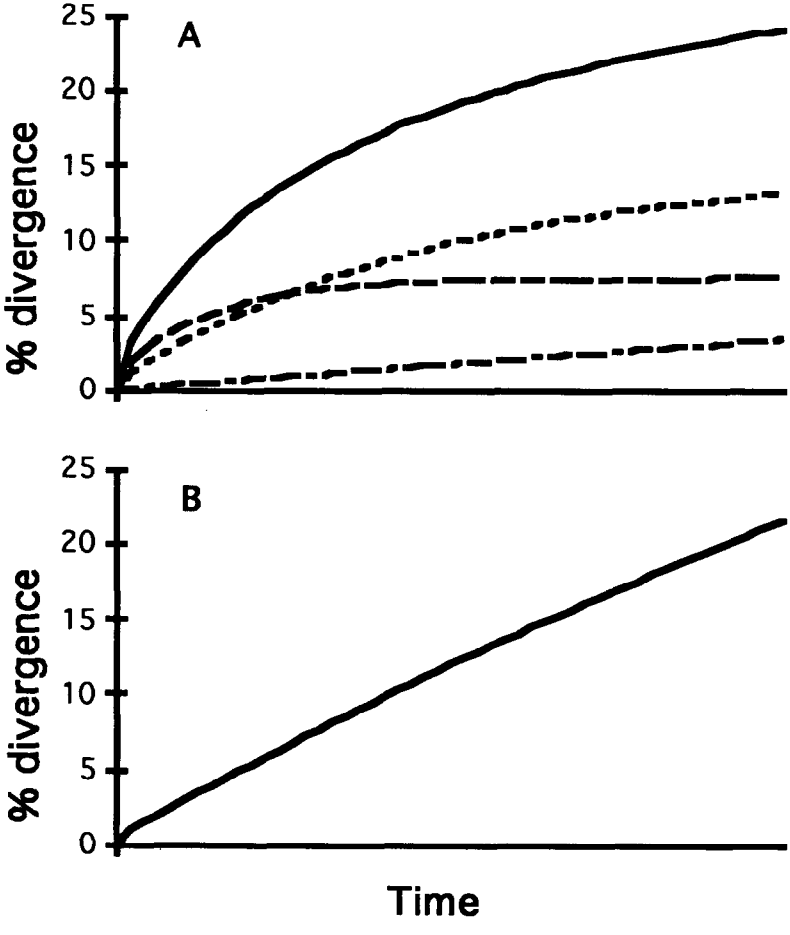

FIG. 8.—Effect of different rate classes on idealized divergence for clocklike molecules. A, Observed divergence curves. The top curve is the divergence of a molecule made up of the three different rate classes. The divergences of the individual classes are shown below the top composite curve. In this model the three classes made up 10%, 20%, and 70% of the sites and evolved at relative rates of 100 ×, 30 ×, and 1, respectively. B, Expected curve for a one-rate-class molecule over the same levels of divergence as in A. In the three-rate-class model, saturation effects dominate the curve shape at low levels of divergence, and linearity occurs only after the faster classes saturate.

Most who have compared rRNA divergence with time have not seen the level of inconstancy that we are reporting (Hasegawa et al. 1985*a*; Ochman and Wilson 1987; Wolfe et al. 1989; Mindell and Honeycut 1990). What is the difference? It could be caused by some unique aspect of the likelihood test, but the relative-rate variation seen (fig. 7) is not unique to likelihood; it also is apparent in pairwise-distance comparisons and in parsimony analysis. Furthermore, likelihood models incorporating a molecular clock have been used, by others (Hasegawa et al. 1985*b*; Li and Tanimura 1987), without yielding gross rate differences. A more plausible explanation is that the inconstancy is correlated with the low divergence levels surveyed in our study. Two other studies of rRNA have focused on low to moderate levels of divergence, and both show significant rate variability. The analysis of 18S rRNAs in *Candida,* by Barnes et al. (1991), covers several low to intermediate ranges of divergence in fungi. Although rate variation was not specifically tested, an examination of their tree reveals two species, *C. lusitaniae,* and *C. krusei,* that occupy intermediate positions and appear to have greatly accelerated rates of evolution relative to their neighboring branches in the tree. Hixson and Brown's (1986) study of mitochondrial 12S rRNA in primates also spans low to moderate levels of divergence. By comparing these rates with those of other mitochondrial genes, they found that divergence was not linear but, instead, appeared to be the sum of a bi- to multicomponent rate.

Langley and Fitch (1974) noted, in protein genes, a correlation between low divergence and high deviation from clocklike behavior. They speculated that, if rates of molecular evolution change, over time, within a bounded range, then one might expect shorter time spans to show greater deviations from the average rate, while longer time spans would tend to average out such deviations. In comparisons of rRNA genes a low divergence level may be more variable for an additional reason. When more distantly related organisms are compared, the most variable portions of the genes become unalignable and are therefore excluded. These exclusions result in fewer classes of sites in the molecule and, perhaps, in a more uniform rate as measured by single-rate-class models. Furthermore, highly variable sites that are not excluded are likely to be saturated in comparisons of distantly related organisms, so that most or all branch lengths would be derived from the more linear portions of the composite rate curve (fig. 8). Taken together, these inferences suggest that, at lower levels of divergence, treating rRNAs as though all sites have an equal probability of change may introduce systematic errors into divergence estimates.

Acknowledgments

We thank Andrea Gargas for the development of the additional primers used for sequencing of the nuclear gene; Joe Felsenstein for sending a prerelease version of his SEQBOOT program and for his advice on the likelihood analysis; and Mary Berbee, Tom White, John Taylor, Walter Fitch, and anonymous reviewers for useful comments on the manuscript. Funding was provided by NSF grant BSR 8700391 to T.D.B.

LITERATURE CITED

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.

BAURA, G., T. M. SZARO, and T. D. BRUNS. 1992. *Gastroboletus laricinus* is a recent derivative of *Suillus grevillei:* molecular evidence. Mycologia **84**:592–597.

BERBEE, M. L., and J. W. TAYLOR. Convergence in ascospore discharge mechanism among the

pyrenomycete fungi based on 18S ribosomal DNA sequence. Mol. Phylogenet. Evol. (accepted).

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitchondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

BRUNS, T. D., R. FOGEL, and J. W. TAYLOR. 1990. Amplification and sequencing of DNA from fungal herbarium specimens. Mycologia **82**:175–184.

BRUNS, T. D., R. FOGEL, T. J. WHITE, and J. D. PALMER. 1989. Accelerated evolution of a false-truffle from a mushroom ancestor. Nature **339**:140–142.

BRUNS, T. D., and J. D. PALMER. 1989. Evolution of mushroom mitochondrial DNA: *Suillus* and related genera. J. Mol. Evol. **28**:349–362.

BRUNS, T. D., T. J. WHITE, and J. W. TAYLOR. 1991. Fungal molecular systematics. Annu. Rev. Ecol. Syst. **22**:525–564.

CAMBARERI, E. B., B. C. JENSEN, E. SCHABTACH, and E. U. SELKER. 1989. Repeat-induced G-C to A-T mutations in *Neurospora.* Science **244**:1571–1575.

CEDERGREN, R., M. W. GRAY, Y. ABEL, and D. SANKOFF. 1988. The evolutionary relationships among known life forms. J. Mol. Evol. **28**:98–112.

CUMMINGS, D. J., J. M. DOMENICO, J. NELSON, and M. L. SOGIN. 1989. DNA sequence, structure, and phylogenetic relationship of the small subunit rRNA coding region of mitochondrial DNA from *Podospora anserina.* J. Mol. Evol. **28**:232–241.

DESALLE, R., T. FREEDMAN, E. M. PRAGER, and A. C. WILSON. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila. J. Mol. Evol.* **26**:157–164.

DESALLE, R. and A. R. TEMPLETON. 1988. Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila.* Evolution **42**:1076–1084.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1985. Confidence intervals on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

———. 1990. PHYLIP—phylogenetic inference package, version 3.3. Computer programs distributed by the author, Department of Genetics, University of Washington, Seattle.

FUKAMI-KOBAYASHI, K., and Y. TANTENO. 1991. Robustness of maximum likelihood tree estimation against different patterns of base substitutions. J. Mol. Evol. **32**:79–91.

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.

GILLESPIE, J. H., and C. H. LANGLEY. 1979. Are evolutionary rates really variable? J. Mol. Evol. **13**:27–34.

GOJOBORI, T., W.-H. LI, and D. GRAUR. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. **18**:360–369.

GRAY, M. W., D. SANKOFF, and R. J. CEDERGREN. 1984. On the evolutionary decent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res. **12**:5837–5852.

GUADET, J., J. JULIEN, J. F. LAFAY, and Y. BRYGOO. 1989. Phylogeny of some *Fusarium* species, as determined by large-subunit rRNA sequence comparison. Mol. Biol. Evol. **6**:227–242.

GUTELL, R. R., B. WEISER, C. R. WOESE, and H. F. NOLLER. 1985. Comparative anatomy of 16-S-like ribosomal RNA. Prog. Nucleic Acids Res. Mol. Biol. **32**:155–216.

GYLLENSTEN, U. B., and H. A. ERHLICH. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA DQα locus. Proc. Natl. Acad. Sci. USA **85**:7652–7658.

HASEGAWA, M., H. HIROHISHA, and T. YANO. 1985*a.* Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HASEGAWA, M., Y. IIDA, T-A. YANO, F. TAKAIWA, and M. IWABUCHI. 1985*b.* Phylogenetic

relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. J. Mol. Evol. **22**:32–38.

HIXSON, J. E., and W. M. BROWN. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. Mol. Biol. Evol. **3**:1–18.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

LANGLEY, C. H., and W. M. FITCH. 1974. An examination of the constancy of the rate of molecular evolution. J. Mol. Evol. **3**:161–177.

LI, W.-H., and M. TANIMURA. 1987. The molecular clock runs more slowly in man than in apes and monkeys. Nature **326**:93–96.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1984. Nonrandomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J. Mol. Evol. **21**: 58–71.

MINDELL, D. P., and R. L. HONEYCUT. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Annu. Rev. Ecol. Syst. **21**:541–566.

OCHMAN, H., and A. C. WILSON. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**:74–86.

PALUMBI, S. R. 1989. Rates of molecular evolution and the fraction of nucleotide positions free to vary. J. Mol. Evol. **29**:180–187.

SMITH, T. L. 1989. Disparate evolution of yeasts and filamentous fungi indicated by phylogenetic analysis of glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci. USA **86**: 7063–7066.

SOGIN, M. L., and J. H. GUNDERSON. 1987. Structural diversity of eukaryotic small subunit ribosomal RNAs: evolutionary implications. Ann. NY Acad. Sci. **503**:125–139.

SWOFFORD, D. L. 1990. PAUP: phylogenetic analysis using parsimony, version 3.0L. Computer program distributed by the Illinois Natural History Survey, Champaign.

VAN DE PEER, Y., J.-M. NEEFS, and R. DE WACHTER. 1990. Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms, and mitochondrial origins. J. Mol. Evol. **30**:463–476.

WHITE, T. J., T. D. BRUNS, S. B. LEE, and J. W. TAYLOR. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315–322 *in* M. A. Innes, D. H. Gelfand, J. Sninsky, and T. W. White, Eds. PCR protocols: a guide to methods and applications. Academic Press, New York.

WILSON, A. C., R. L. CANN, S. M. CARR, M. GEORGE, U. B. GYLLENSTEN, K. M. HELM-BYCHOWSKI, R. G. HIGUCHI, S. R. PALUMBI, E. M. PRAGER, R. D. SAGE, and M. STONEKING. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linnaean Soc. **26**:375–400.

WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.

WOLFE, K. H., M. GOUY, Y.-W. YANG, P. M. SHARP, and W.-H. LI. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86**:6201–6205.

WOLFE, K. H., W.-H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**: 9054–9058.

WESLEY M. BROWN, reviewing editor

Received July 14, 1991; revision received April 8, 1992

Accepted April 10, 1992