Unusually High Evolutionary Rate of the Elongation Factor 1α Genes from the Ciliophora and Its Impact on the Phylogeny of Eukaryotes

David Moreira, Hervé Le Guyader, and Hervé Philippe

Laboratoire de Biologie Cellulaire, Unité de Recherche Associée 2227, Centre National de la Recherche Scientifique, Université Paris–Sud, Orsay, France

The elongation factor 1α (EF-1α) has become widely employed as a phylogenetic marker for studying eukaryotic evolution. However, a disturbing problem, the artificial polyphyly of ciliates, is always observed. It has been suggested that the addition of new sequences will help to circumvent this problem. Thus, we have determined 15 new ciliate EF-1α sequences, providing for a more comprehensive taxonomic sampling of this phylum. These sequences have been analyzed together with a representation of eukaryotic sequences using distance-, parsimony-, and likelihood-based phylogenetic methods. Such analyses again failed to recover the monophyly of Ciliophora. A study of the substitution rate showed that ciliate EF-1α genes exhibit a high evolutionary rate, produced in part by an increased number of variable positions. This acceleration could be related to alterations of the accessory functions acquired by this protein, likely to those involving interactions with the cytoskeleton, which is very modified in the Ciliophora. The high evolutionary rate of these sequences leads to an artificial basal emergence of some ciliates in the eukaryotic tree by effecting a long-branch attraction artifact that produces an asymmetric topology for the basal region of the tree. The use of a maximum-likelihood phylogenetic method (which is less sensitive to long-branch attraction) and the addition of sequences to break long branches allow retrieval of more symmetric topologies, which suggests that the asymmetric part of the tree is most likely artifactual. Therefore, the sole reliable part of the tree appears to correspond to the apical symmetric region. These kinds of observations suggest that the general eukaryotic evolution might have consisted of a massive radiation followed by an increase in the evolutionary rates of certain groups that emerge artificially as early branches in the asymmetric base of the tree. Ciliates in the case of the EF-1α genes would offer clear evidence for this hypothesis.

Introduction

The advent of molecular phylogeny has forever changed our understanding of protist evolution by providing new tools and a phylogenetic framework based on criteria more objective than the analysis of structural characters. Initial work was carried out using the small-subunit ribosomal RNA (SSU rRNA) as a phylogenetic marker. The eukaryotic phylogeny that emerged from SSU rRNA sequence comparison consisted of a basal region with several protist groups emerging in a stepwise fashion and a large apical unresolved radiation (known as the “crown”) grouping together plants, animals, fungi, and several protist phyla (Sogin 1991). Nevertheless, the rRNA-based phylogenies were soon revealed to be sensitive to problems derived from drastic differences in GC content among taxa (Loomis and Smith 1990; Galtier and Gouy 1995). As a consequence, since they are not affected by this problem, several proteins such as actin, α- and β-tubulin, glyceraldehyde-3-phosphate dehydrogenase, and the protein elongation factor 1α (EF-1α) are currently used as alternative phylogenetic markers.

In particular, EF-1α has become widely employed. This protein plays a central role in protein synthesis, forming a ternary complex with aminoacyl-tRNA and GTP (Yager and von Hippel 1987). In addition, it interacts with some cytoskeletal proteins, especially actin (Durso and Cyr 1994; Shiina et al. 1994). Some relevant phylogenetic questions have been successfully addressed using this marker, including the phylogenies of some groups of metazoans (Kobayashi, Wada, and Satoh 1996; McHugh 1997), the sisterhood of the fungal and animal clades (Baldauf and Palmer 1993; Hasegawa et al. 1993), and the monophyly of the slime molds and their inclusion within the crown of the eukaryotic tree (Baldauf and Doolittle 1997). However, it was quite disturbing that ciliates included so far in several analyses (Tetrahymena pyriformis, Euplotes crassus, and Stylonychia lemmiae) did not form a monophyletic clade in any EF-1α tree (Adoutte et al. 1996; Baldauf and Doolittle 1997; Yamamoto et al. 1997). This seriously contradicts a large amount of morphological, developmental, and molecular data that clearly support the monophyly of the phylum Ciliophora (for review, see Schlegel and Eiser 1996). As a result, the hypothetical ciliate polyphyly derived from EF-1α analyses has been regarded as artifactual. Several arguments have been put forward to explain this apparent contradiction. First, only a few ciliate sequences, ill representing the diversity of this phylum, were available (Baldauf and Doolittle 1997). Second, the ciliates have a faster rate of sequence divergence, which would generate a long-branch attraction phenomenon (Philippe and Adoutte 1996; Adoutte et al. 1996; Baldauf and Doolittle 1997; Roger et al. 1999).

Indeed, long-branch attraction (Felsenstein 1978) is a common artifact affecting phylogenetic reconstruction. When a distant outgroup is used, fast-evolving taxa are attracted by the long branch of the outgroup, and the resulting trees display artifactual asymmetric bases. Based on this idea, it has been proposed that eukaryotic evolution might have consisted of a massive radiation followed by an increase in the evolutionary rates of certain groups that emerge artificially as early branches,
producing an asymmetric tree base (Philippe and Adoutte 1998). In addition, since the groups exhibiting higher evolutionary rates can be different for different phylogenetic markers, this model also helps to explain the incongruences observed among different phylogenetic markers for the likely artifactual asymmetric regions of the trees.

The EF-1α-based phylogeny displays an asymmetrical base, and the artifactual polyphyly of ciliates, located in this region, is an interesting case study. Addition of taxa to break long branches was suggested as a means to alleviate the long-branch attraction artifact (Hendy and Penny 1989), and therefore, if the model explaining the asymmetric bases is correct, the addition of new ciliate sequences to the EF-1α phylogeny must lead to a more symmetric tree topology.

We thus determined 15 new ciliate sequences, including representatives of their main classes. The new sequences, together with a selection of published eukaryotic EF-1α sequences, were analyzed using distance-, parsimony-, and likelihood-based phylogenetic methods. These analyses still failed to recover the monophyly of Ciliophora, with an artifactual basal emergence of the fast-evolving species. Nevertheless, the comparison of different data sets has shown that, as expected, the tree becomes increasingly symmetrical with the addition of new sequences, supporting the model described above. The unusually high evolutionary rate of the ciliate EF-1α genes is in part produced by an increased number of variable positions. Their particular mode of evolution is difficult to explain by modifications in the translational activity. Conversely, it could be related to alterations of the accessory functions involving interactions with the cytoskeleton, which has undergone important modifications in the Ciliophora.

Materials and Methods
Strains and Growth Conditions

Species from seven different classes of ciliates were used in this study. Cells of *Spathidium* sp. (class Litostomata) and of *Kentrophoros* sp. (class Karyorelictea) were kindly provided by Dr. A. Baroin-Tourancheau. The rest of the species were grown on liquid media containing the following nutrients: 1.5% bacto-proteose peptone (Difco) for *T. pyriformis* (class Oligohymenophorea); infusion of the marine green alga *Lactuca* sp. and living cells of the bacterium *Aerobacter aerogenes* for *Paranophrys carnivora*, *Telotrichodium henneguyi* (class Oligohymenophorea), and *Colpoda inflata* (class Colpodea); living cells of the cyanobacterium *Phormidium autumnale* for *Naxella* sp. (class Nassophorea); living cells of the ciliate *T. pyriformis* for *Euplotes aediculatus* and *Stylonychia mytilus* (class Hypotrichia); wheat and rice infusion for *Blepharisma japonicum* (class Heterotricha); and living cells of the cryptophyte alga *Cryptomonas* sp. for *Stentor coerulescens* (class Heterotricha). All the strains except *Naxella* sp. (found in a pond in Tivoli, Italy) were isolated from different ponds and soil samples from the campus of the Paris XI University (Orsay, France). Gene Amplification, Cloning, and Sequencing

Ciliate genomic DNA was isolated using a hexadecyltrimethylammoniumbromide (CTAB) purification method as previously described in Winnepenninckx, Backeljau, and de Wachter (1993). Almost complete EF-1α coding sequences (around 1,270 bp) were amplified with different combinations of the forward primer 1F and the reverse primers 10R1 and 10R2, described by Baldauf and Doolittle (1997), and a new forward primer 1A (GGT ATT GGA CAY GTY GAY TCN GG), corresponding to the peptide VIGHVDSG. Partial regions from the EF-1α-coding sequences (350–700 bp) were amplified with the above described primers, together with the forward primers 3F and 4F2 and the reverse primers 4R and 7R (Baldauf and Doolittle 1997). These partial amplicons were screened to study the possible presence of introns, since during PCR amplifications of complete EF-1α-coding regions, the smaller, intronless loci might have been preferentially amplified. PCR amplifications were carried out as described by Baldauf and Doolittle (1997).

After each amplification and cloning assay, plasmid DNAs from 30 clones containing inserts of the expected size were analyzed by endonuclease restriction with the enzymes *EcoRI*, *HindIII*, *SalI*, and *Sau3AI* (New England Biolabs). In some cases, the coexistence of two bands with different restriction patterns, which were subsequently shown to be duplicated EF-1α genes, was detected. Plasmid DNAs were sequenced using the Thermo Sequenase premixed cycle sequencing kit (Amersham). For each amplification product (including all those differentiated by their restriction profiles), a minimum of 10 clones were analyzed by partial sequencing (around 1,000 nt) in order to detect possible different clones not distinguishable by the restriction analysis. Sequences were obtained automatically with a Vistra DNA Sequencer 725 (Amersham). Both DNA strands of one representative of each distinct clone, as identified by the restriction analysis or by the partial sequencing, were completely sequenced, as was one strand of another clone to check for possible Taq polymerase errors. When discrepancies were observed, they were resolved by sequencing of additional clones. Sequences have been submitted to GenBank under the accession numbers AF056096–AF056109.

Phylogenetic Analysis

Assembly and translation of the ciliate sequences were carried out using the programs of the GCG Wisconsin Package, version 9.1 (Genetics Computer Group, Madison, Wis.). All the eukaryotic EF-1α sequences available in data banks were identified using a BLAST search mail facility (blast@ncbi.nlm.nih.gov). The blast2retp and retp2ali programs (P. Lopez, personal communication) allowed the automatic retrieval of these sequences and their writing into a MUST-compatible file. Partial sequences or those containing likely sequencing errors were discarded, leading to a final number of 197 eukaryotic sequences. Thirteen archaecal EF-1α sequences were included as the outgroup. All of these sequences were manually aligned using the ED

Different sequence subsets (see Results) were aligned in order to study the impact of the addition of new sequences on the EF-1α phylogeny. Well-known monophyletic groups (fungi, metazoans, and angiosperms) were represented by small but representative numbers of sequences to have data sets of limited size to facilitate the time-consuming analyses. Additionally, EF-1α sequences known to be under developmental regulation or low expression control were removed (e.g., Xenopus laevis thersaun, Strongylocentrotus purpuratus, Porphyra purpurea tef-s). On the other hand, almost all available protist sequences were used. The most important exception was the sequence from the hypotrich ciliate S. lemnae, which was eliminated to reduce the number of taxa, considering that this genus was well represented by the species S. mytilus (97.5% identity with S. lemnae). The sequence from the microsporidian Glugea plecoglossi (Kamaishi et al. 1996) was also excluded from the analysis, since it is extremely divergent and likely to be misplaced in phylogenetic trees, taking into account the compelling evidence to consider microsporidians very derived fungi (Müller 1997).

Phylogenetic trees were constructed using distance (neighbor-joining [NJ]) (Saitou and Nei 1987), maximum-parisimony (MP), and maximum-likelihood (ML) methods with the MUST version 1.0 package (program NJ) (Philippe 1993), PAUP version 3.1 (Swofford 1993), and PROML version 2.3 (Adachi and Hasegawa 1996), respectively. Phylogenetic distances were computed with the method of Kimura (1983). MP trees were obtained from 100 random addition heuristic search replicates and the tree bisection-reconnection (TBR) branch-swapping option. ML trees were constructed by the quick-add operational taxonomic units search employing the JTT model of amino acid substitution and the tree bisection-reconnection (TBR) branch-swapping option. ML trees were computed with the method of Kimura (1983). MP trees were obtained from 100 random addition heuristic search replicates and the tree bisection-reconnection (TBR) branch-swapping option. ML trees were constructed by the quick-add operational taxonomic units search employing the JTT model of amino acid substitution and the tree bisection-reconnection (TBR) branch-swapping option.

Mutational saturation of the EF-1α sequences from a given taxonomic group was measured by comparison of the number of substitutions inferred by MP or ML methods with the number of differences observed for each pair of sequences (Philippe et al. 1994).

Analysis of Sequence Variability

To analyze the number of substitutions of the ciliate EF-1α sequences and compare it with those of other taxonomic groups, we distributed 162 eukaryotic and 13 archaean aligned sequences into 8 different groups. Ciliates were analyzed as a single group (“all Ciliophora”) as well as being decomposed into the sequences from the genus Euplotes (“Euplotes spp.”) and those from the rest of the ciliates (“other Ciliophora”). Thus, groups were defined as follows: Metazoa (118 sequences), Fungi (20 sequences), Magnoliophyta (angiosperms) (14 sequences), all Ciliophora (19 sequences), Euplotes spp. (4 sequences), other Ciliophora (15 sequences), Diplomonadida (5 sequences), and Archaea (13 sequences). The most parsimonious tree was computed for each group using PAUP, version 3.1, with 10 random additions of taxa and the TBR branch-swapping option. The number of variable positions as well as the number of substitutions for each sequence position within each group was thus inferred. The numbers of variable positions for the same taxonomic groups were also calculated using SSU RNA and α-tubulin sequence data for comparison with the EF-1α data.

Similarly, the number of substitutions for each position of the EF-1 alignment was inferred and displayed on an alignment of the consensus sequences of the different taxonomic groups. Constant positions were represented by their amino acid letters, those undergoing one or two substitutions were replaced by light gray boxes, while those that have undergone three or more substitutions were replaced by dark gray boxes.

Results

Characterization of Ciliate EF-1α Sequences

PCR products of about 1,270 bp amplified from ciliate genomic DNAs using the primer pairs 1F-10R1, 1F-10R2, 1A-10R1, and 1A-10R2 were cloned and screened by restriction analysis (30 clones for each amplification) and partial sequencing. The EcoRI restriction analysis proved to be especially suitable for differentiation of multiple tef genes within a given species. Thus, two different tef loci (showing different EcoRI restriction patterns) were detected in the species B. japonicum, E. eucalitrus, Spathidium sp., and S. coerulues, while only single tef loci were identified in the rest of the species (C. inflata, Kentrophoros sp., Naxella sp., P. carnivora, P. tetraurelia, S. mytilus, and T. henneyguyi). Fifteen new EF-1α genes were cloned and sequenced. None of these contained introns in the region covered by the primers used. Failure to detect introns could have been a result of PCR bias, since the reactions designed to amplify large DNA fragments could have led to the recovery of only the smaller, intronless loci. To eliminate this possibility, we additionally amplified several
specific small fragments covering almost the complete EF-1α-coding genes for all the tested species (see Materials and Methods). Sequencing of these fragments confirmed a lack of introns. The relatively high number of amplifications and clones screened supported the theory that, indeed, ciliate *tef* loci, at least in these species, are intronless.

**EF-1α-Based Ciliate Phylogeny**

A phylogenetic analysis of the ciliate EF-1α sequences rooted with apicomplexan sequences (Fig. 1), using NJ, MP, and ML methods, was carried out. The phylogenetic distribution of the duplicated *tef* sequences clearly indicates that several independent duplications took place in ciliate *tef* gene evolution, instead of one ancestral event followed by some losses. Thus, each pair of duplicated sequences from the heterotrichs *B. japonicum* and *S. coeruleus* and from the litostomate *Spathidium* sp. appears as a monophyletic group. This strongly suggests that the divergence of these species preceded the duplication of their *tef* loci. In contrast, the duplicated sequences from the two *Euplotes* species revealed a more complex situation. Their observed phylogeny could be explained by an ancestral duplication of the *tef* locus before the divergence of the two species, followed by the loss of one of the *E. aediculatus* alleles and further duplication of its remaining copy. An alternative explanation might be two independent duplications hidden by the high divergence degree of the second *E. crassus* sequence, which leads to its emergence at the base of the *Euplotes* clade by a long-branch attraction artifact. At present, the available information does not favor either hypothesis.

The ML tree did not recover the monophyly of two ciliate classes, oligohymenophoreans and hypotrichs, even when the first 5,000 trees, ordered by their likelihood values, were analyzed. Only the monophyly of heterotrichs (*B. japonicum* and *S. coeruleus*) was retrieved, supported by a high BP value (100) (Fig. 1). Consequently, the topology of our EF-1α-based tree was incongruent with results of rRNA sequence analysis, which support the monophyly of all of these classes (Baroin-Tourancheau et al. 1995; Wright and Lynn 1997). Furthermore, depending on the reconstruction method applied (NJ, MP, or ML), ciliates showed very different relationships, except for the monophyly of heterotrichs, the two sequences from *Spathidium* sp., and those from the genus *Euplotes* (data not shown). The bootstrap support for the deep nodes sustaining these diverse topologies was very low (BP ≤ 45). Altogether, these results showed that EF-1α was not a reliable marker for inferring ciliate phylogeny.

The aberrant ciliate phylogeny may be due to the accumulation of multiple substitutions, which erases part of the phylogenetic signal. In fact, ciliate sequences became saturated above a value of approximately 90 observed differences, reaching a plateau around 130 observed differences, whereas the inferred substitutions (calculated by MP) reach a value as high as 336 (Fig. 2A). This saturation level has an order of magnitude comparable to that of the complete eukaryotic and archaean data set, which displays a maximum number of observed differences of nearly 200, almost three times as small as the maximum number of inferred substitutions (560) (Fig. 2B). The mutational saturation becomes especially critical when sequences show unequal evolutionary rates, as in the case of ciliates. In fact, some ciliate sequences, especially those from *Spathidium* and *Euplotes* species, exhibit branches longer than the rest, which indicates important differences in evolutionary rates. The combination of both factors, mutational sat-
FIG. 2.—Saturation curves for the ciliate (A) and eukaryote plus archaea (B) data sets. On the y axis, the numbers of observed differences between pairs of sequences are shown; on the x axis, the numbers of substitutions between the same two sequences inferred by the MP method are shown. Each dot thus represents one pair of sequences, considering the observed over the inferred number of substitutions. The diagonal represents the ideal case of no more than one substitution per sequence position. Open circles (B) correspond to eukaryotic versus archaean sequence comparisons.

Phylogenetic Analysis of the Expanded Eukaryotic EF-1α Data Set

To test the model explaining the asymmetric base of the eukaryotic EF-1α tree, four different data sets (A, B, C, and D) were analyzed by all three methods. Since ciliate sequences exhibit notable unequal evolutionary rates, only the ML analysis, which is less affected by this factor (Hasegawa and Fujiwara 1993), will be described. The initial data set studied (A) contained a limited sampling of sequences (20 eukaryotic sequences plus 8 archaean sequences as the outgroup) to represent the original situation prior to our studies. The phylogenetic analysis led to a polyphyly of ciliates (fig. 3A), which displayed four independent offshoots as previously reported (Hashimoto and Hasegawa 1996). Three of the ciliate sequences branched out of the crown, and the second E. crassus EF-1α sequence was even the first offshoot in the NJ tree (earlier than the Giardia lamblia sequence).

The addition of our new ciliate sequences (fig. 3B) or of several new protist sequences recently determined by other groups (fig. 3C) again failed to recover the monophyly of ciliates. Tree B again separated ciliates into four independent groups, while tree C (with a smaller ciliate sampling) produced two ciliate assemblages. Interestingly, the addition of several nonciliate sequences to the initial data set led to the recovery of the monophyly of three ciliate sequences (T. pyriformis and both E. crassus sequences), although with very low support (BP = 22).

Finally, all the sequences combined into a fourth data set (D) generated a tree topology in which ciliates did not form a monophyletic clade but were distributed as three discrete groups (fig. 3D), regardless of the reconstruction method employed. A large ciliate group comprised 14 of the 18 ciliate sequences, although their monophyly was very weakly supported (BP = 3). In addition, the apicomplexans P. falciparum and C. parvum and the amoeba E. histolytica branch artifactually within this group. Moreover, in the NJ and MP trees, all the Euplotes sequences emerged out of this ciliate group, with the most divergent E. crassus sequence being the earliest offshoot of the eukaryotic tree (data not shown). The two Spathidium sp. sequences were also located outside of the main ciliate clade with a very unstable position, varying from a basal emergence together with the diplomonads and the trichomonads in the MP tree (with a very low BP of 8) to a more central region after the emergence of these phyla in the NJ and ML trees. All these observations reflect the extreme instability of the phylogenetic positions of the Euplotes and Spathidium sequences, a result possibly related to long-branch attraction artifacts stemming from their high evolutionary rates. Finally, the ML tree also displayed the hypotrich S. mytilus and the karyorelictean Kentrophoros sp. outside of the main ciliate group, as the first offshoot next to a Euglenozoa-Magnoliophyta clade.

A positive effect of sequence addition was detected through the evolution of the tree symmetry. This parameter is expressed with an index (Iₚ) that ranges from 0, for a completely symmetric tree, to 1, for a completely asymmetric tree (Heard 1992). All the trees obtained with the initial data set were highly asymmetric (Iₚ of 0.73) (table 2), especially in their basal regions. Interestingly, the increase in sequence number within the two intermediate data sets with ciliate (data set B) and nonciliate species (data set C) strongly affected the symmetry of the new trees (Iₚ values of 0.33 and 0.39, respectively). Finally, the most complete data set (D) showed very low tree asymmetry (Iₚ = 0.18). We estimated the net gain of symmetry by calculating the Iₚ of
Fig. 3.—The impact of sequence addition. ML trees were constructed using data sets A (initial data set), B (addition of ciliate sequences), C (addition of diverse protist sequences), and D (combination of data sets B and C). Ciliate species are shown in bold. Numbers at each node are bootstrap values. Scale bars represent 5 substitutions per 100 positions.
the 20 species from the initial data set within the new tree topologies obtained with the other three data sets (subset A in table 2). The $I_p$ for subset A, with an initial value of 0.73 (i.e., that of tree A), decreased to 0.56 and 0.55 in trees B and C, respectively, and reached 0.45 in the tree D. Comparison of the different reconstruction methods showed that the ML retrieved the most symmetric trees from almost all data sets, while the NJ and MP methods did not show important differences between them (table 2). The observed increase of $I_p$ is thus in good agreement with the predictions of the model.

The improvement of the inferred phylogeny is also well manifested by the fact that in the final tree, the monophyly of ciliates requires a difference of likelihood of only $-25.5$ (representing an SE of 0.59) against the most likely tree (with a likelihood of $-17.757$), which is statistically insignificant (Kishino and Hasegawa 1989). A tree with monophyletic ciliates corresponds to 3,041 steps, compared with 3,031 for the most parasimonious tree. Moreover, the monophyly of alveolates (i.e., the ciliates plus the two apicomplexans *P. falciparum* and *C. parvum*) was recovered with the same small increase of 10 steps (table 1). The taxonomic sampling had an interesting effect on these values, since in data set A, the monophyly of the four ciliate sequences requires an increase of 15 steps (1,925 instead of 1,910 for the most parasimonious tree), similar to the values obtained using data sets B and C (14 and 16 steps, respectively) (table 1). However, these values, expressed as percentages of the respective total numbers of steps, are much smaller for data sets B and C (0.55 and 0.67, respectively) than for data set A (0.79). As expected, the percentage is still smaller for data set D (0.33). Alveolate monophyly requires similar values (table 1). Notably, the monophyly of a smaller number of ciliate sequences (4 in data set A) is more difficult to retrieve than that of a larger number (18 in data sets B and D), which demonstrates the positive effect of sequence addition.

These results are related to the fact that statistical support for the different nodes of the tree varies depending on the addition of new sequences. Actually, the tree derived from the smallest data set (A) showed three different regions. A first region comprising the earliest branches supported by BPs above 50, a central region with BPs ranging from 39 to 48 (except for the BP of 21 for the branching of the Magnoliophyta), and a third region containing some taxa of the crown supported by very high BPs (above 98). The addition of new taxa to the analysis (trees B and C) did not alter this regionalization of BP values, although these values strongly decreased in the central region (approximately from 39–48 to 10–40). Using data set D, this decrease was accentuated (central-region BP values were between 3 and 23 (fig. 3). Interestingly, BP values in the other two regions of the tree, especially those of the crown, did not experience pronounced changes.

The monophyly of Diplomonadida, Euglenozoa, Fungi, Metazoa, and Mycetozoa (including dictyostelids, myxogastriads, and protostelids) (BPs close to 100), the early emergence of diplomonads (*Giardia* and relatives) (BPs of 57–91) and of the parabasalian *Trichomonas* (BPs of 71–88), and the sisterhood of Fungi and Metazoa (BPs of 96–100) were well supported by all of the data sets and methods employed. However, the sister group of this last clade could not be reliably determined with the available data. Depending on the taxonomic sampling and on the method used, the Mycetozoa or the rhodophyte *P. purpurea* is the closest group to the animal-fungal clade (data not shown). Interestingly, a monophyletic group containing the Euglenozoa and the Magnoliophyta (as first noticed by Roger et al. 1999) was observed in all the trees after the addition of new species (ciliates or other protists), with BPs ranging from 40 to 68.

### Table 1

<table>
<thead>
<tr>
<th>Topology</th>
<th>A (28 species)</th>
<th>B (42 species)</th>
<th>C (37 species)</th>
<th>D (51 species)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ciliate monophyly</td>
<td>(D = 15) (0.79%)</td>
<td>(D = 14) (0.55%)</td>
<td>(D = 16) (0.67%)</td>
<td>(D = 10) (0.33%)</td>
</tr>
<tr>
<td>(\Delta \text{LnL} = 0.76\ SE)</td>
<td>(\Delta \text{LnL} = 2.52\ SE)</td>
<td>(\Delta \text{LnL} = 0.91\ SE)</td>
<td>(\Delta \text{LnL} = 0.59\ SE)</td>
<td></td>
</tr>
<tr>
<td>Alveolate monophyly</td>
<td>(D = 17) (0.89%)</td>
<td>(D = 15) (0.59%)</td>
<td>(D = 16) (0.67%)</td>
<td>(D = 10) (0.33%)</td>
</tr>
<tr>
<td>(\Delta \text{LnL} = 0.88\ SE)</td>
<td>(\Delta \text{LnL} = 1.38\ SE)</td>
<td>(\Delta \text{LnL} = 0.91\ SE)</td>
<td>(\Delta \text{LnL} = 0.19\ SE)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Recon. Method</th>
<th>ML</th>
<th>MP</th>
<th>NJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Taxa</td>
<td>All Taxa</td>
<td>All Taxa</td>
<td>All Taxa</td>
<td></td>
</tr>
<tr>
<td>Subset A</td>
<td>Subset A</td>
<td>Subset A</td>
<td>Subset A</td>
<td></td>
</tr>
<tr>
<td>A ....</td>
<td>0.73</td>
<td>0.73</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>B ....</td>
<td>0.33</td>
<td>0.56</td>
<td>0.39</td>
<td>0.69</td>
</tr>
<tr>
<td>C ....</td>
<td>0.39</td>
<td>0.55</td>
<td>0.26</td>
<td>0.50</td>
</tr>
<tr>
<td>D ....</td>
<td>0.18</td>
<td>0.45</td>
<td>0.25</td>
<td>0.65</td>
</tr>
</tbody>
</table>

**Note.** — *D* is the difference in the numbers of steps, and \(\Delta \text{LnL}\) is the difference in likelihoods between the best unconstrained tree and trees constraining ciliate or alveolate monophyly. Numbers in parentheses are the percentages of the numbers of steps with regard to the shortest tree.

**Analysis of Variation Rates**

We determined the numbers of variable positions for eight taxonomic groups (Metazoa, Fungi, Magnoliophyta, all Ciliophora, genus *Euplotes*, the rest of Ciliophora, Diplomonadida, and Archaea) and three different genes (EF-1α, SSU rRNA, and α-tubulin) (table 3).
Among the eukaryotic EF-1α sequences, ciliates exhibited the highest number of variable positions (251 out of 387 aligned positions, twice as many as the taxonomically diverse fungi). This large proportion of variable positions indicates that, during the evolutionary history of this group, a large fraction of amino acid positions were free to vary, i.e., a large number of covarions existed (Fitch and Markowitz 1970). This number was comparable to that of the Archaea (276), which represents a taxonomic domain as broad as the whole eukaryotic one. Interestingly, the Diplomonadida, a protist group not branching within the terminal crown, with only five EF-1α sequences, showed a high number of variable positions (148).

In contrast, ciliates did not show high variability for the other two markers, which were analyzed using similar taxonomic samplings (table 3). They possess 237 variable positions for SSU RNA, similar to Metazoa (272) and Fungi (204). On the other hand, their α-tubulin is less variable (65 variable positions) than are those of Metazoa (145) and Fungi (224). An important consequence of the great ciliate EF-1α variability is the absence of specific signatures for this group. In fact, signatures are the main reason that the monophyly of groups is recovered even when the sequences show a high degree of saturation (Lopez, Forterre, and Philippe 1998). A good example is provided by the diplomonads, which exhibit a high number of variable positions (148) but also a remarkable number of signature positions (18) that are different or variable in the other groups. This fact leads to very high support (a BP of 100 for all of the methods) for the monophyly of this group.

Additionally, we studied the distribution of the variable positions along the EF-1α primary structures to verify whether some known protein motifs are affected by the unusual variability of the ciliate sequences. For this purpose, the inferred number of substitutions for each position was displayed on an alignment of the consensus sequences of these groups (fig. 4). Notably, some EF-1α motifs involved in the translational activity, such as the GDP-, GTP-, and tRNA-binding regions, are well conserved in Ciliophora, while high variability is concentrated in other regions. Among these motifs, some putative motifs of interaction with actin are most interesting.

### Discussion

**Ciliate tef Gene Evolution**

The analysis of the 15 new ciliate tef genes analyzed here showed three interesting characteristics: their lack of introns, the fact that multiple tef copies in different ciliate species appear to be the result of several independent duplications, and strong unequal evolutionary rates. Intron absence might have been the result of a bias in the PCR amplifications, but we tried to diminish this possibility as much as possible and consistently found no introns. The fact that tef genes directly cloned from genomic DNA of *E. crassus* (Bergemann et al. 1996) and *S. lemnae* (Bierbaum, Donhoff, and Klein 1991) also lacked introns strongly supports actual intron absence. On the other hand, the existence of more than one tef locus per genome is very frequent among eukaryotes, and evidence exists for multiple independent duplication events within different taxonomic groups.

At first sight, the high variability of the ciliate EF-1α sequences with regard to their strong differences of evolutionary rate seemed not to affect the functional domains of the molecule. Indeed, even the most divergent of these sequences show a general high degree of conservation of all of the characterized motifs involved in the translational activity, such as the GDP-, GTP-, and tRNA-binding regions (fig. 4). Other ciliate molecules playing a major role in the translation process, such as rRNAs (Van de Peer, Van der Auwera, and de Wachter 1996), do not display an increased evolutionary rate equivalent to that exhibited by the EF-1α sequences (see table 3). Both observations suggest that the translational activity of ciliates has not undergone a very divergent evolution. However, it has recently been demonstrated that the EF-1α plays additional roles in eukaryotic cells. Compelling evidence exists of the EF-1α interaction with cytoskeletal proteins, particularly actin (Demma et al. 1990; Itano and Hatano 1991), but also tubulin (Shi-...
F I G . 4.—Position-by-position analysis of sequence variability. Consensus sequences were obtained for eight taxonomic groups, and the number of substitutions for each position within each group, inferred by the MP method, is represented. Amino-acid letters correspond to constant positions, light gray boxes correspond to positions that have undergone one or two substitutions, and dark gray boxes correspond to positions that have undergone three or more substitutions. GTP- (bold letters), tRNA-, and actin- (A-b) binding domains are indicated.

Moreover, EF-1α has been found to promote the bundling of actin filaments and the severing of microtubules (Demma et al. 1990; Shiina et al. 1994). Such activities have also been described for protists, including ciliates (Kurasawa et al. 1996).

Interestingly, the overall actin-based phylogeny of eukaryotes is congruent in some respects with that based on EF-1α sequences. It shows a similar polyphyly of ciliates and a very early emergence of some sequences, especially those from the genus Euplotes (Bhattacharya and Weber 1997; Philippe and Adoutte 1998). Actin is a quantitatively minor protein in ciliates, and its function appears to be limited to food vacuole formation (Cohen, Garreau de Loubresse, and Beisson 1984). This limitation implies a relaxation on its selective constraints that could explain its accelerated evolutionary rate (Philippe and Adoutte 1998). Hence, the acceleration of the actin mutational rate might have been accompanied by a co-evolutionary acceleration of the mutational rate of the involved EF-1α interacting regions. These regions are not well characterized, though they have been tentatively circumscribed to some short motifs within the amino-terminal and central domains of the protein (Edmons, Murray, and Condeelis 1995). For almost all these of putative motifs, the highest variability among the groups analyzed is seen in ciliates (fig. 4). Altogether, these results could point to a co-evolutionary acceleration of ciliate actin and EF-1α, or even to a loss of their interaction. However, more data, especially from biochemical approaches, will be needed to test this hypothesis.

The Evolutionary Position of Ciliates and Their Impact on the Eukaryotic EF-1α-Based Phylogeny

The phylogenetic analysis of the enhanced EF-1α data set including 15 new ciliate sequences still displayed an artifactual polyphyletic distribution of this phylum in the eukaryotic tree. At first sight, this seems to contradict the idea that the increase of the number of sequences should improve phylogenies (Lecointre et al. 1993; Graybeal 1998). In fact, the tree topologies obtained from our different data sets suggest that the branching of several ciliate sequences outside of the main ciliate group is due to a long-branch attraction phenomenon. This artifact particularly affects the Euplotes sequences, since the NJ and MP methods do not place them within the main ciliate group, displaying these sequences as very early branches. Thus, the order of emergence of the ciliate species correlates roughly with the branch lengths inferred from the analysis of the internal ciliate phylogeny (fig. 1) and contributes to the increased asymmetry of the basal region of the EF-1α tree.

The simple increase in the number of sequences may not overcome the long-branch attraction problems; the addition of slow-evolving sequences is the most effective approach to reduce problems of phylogenetic inconsistency (Kim 1996).
fast-evolving sequences, such as those of the *Euplotes* and *Spathidium* species, did not solve the long-branch attraction problem for ciliates in the NJ and MP analyses. Thus, although the attraction of the *E. crassus* branches toward the tree base was notably attenuated after addition of the *E. aediculatus* sequences, they join the rest of ciliates only in the ML analysis. Therefore, it seems that the choice of adequate sequences could be an even more decisive factor than the simple increase in the amount of sequences (Aguinaldo et al. 1997). Obviously, the lack of a priori criteria to foresee the quality of any given sequence constitutes a problem. Hence, in practice, the sole strategy with which to approach problematic phylogenies is to increase the number of sequences, with the hope that this will also provide sequences of good phylogenetic quality, i.e., slow-evolving ones. In addition, the careful selection of sequences, taking into account only the slow-evolving sequences, can be of great help.

Although the addition of new sequences has not resolved the problem of long-branch attraction for the EF-1a data set, the behavior of the tree symmetry, an important parameter in describing the tree topologies, appears to confirm the assumption of our model of eukaryotic EF-1a evolution. In fact, the addition of new sequences led to a strong increase in the global tree symmetry, from an initial $I_s$ value of 0.73 (data set A) to a final one of 0.18 (data set D) (table 2). A part of this gain of symmetry comes from the simple increase of species within several clades. The most important gain of symmetry with the model is due to the joining of some basal branches emerging asymmetrically in the first tree (data set A) to monophyletic groups in the final tree (data set D). The increase of the sequence number leads from an initial $I_s$ value of 0.73 to a final one of 0.45 for species subset A. This net gain of symmetry due to the breaking of some long branches supports the notion that the asymmetric base of the EF-1a tree is highly artifactual and probably derived from long-branch attraction artifacts. The comparison of the topologies derived from the different reconstruction methods lends additional evidence for this conclusion. In fact, the ML trees displayed the highest symmetry values in almost all cases (table 2). Since the ML method is the least sensitive to long-branch attraction (Kishino and Hasegawa 1989), the fact that the ML tree topologies are more symmetric also allows establishment of a likely correlation between asymmetry and artifactual reconstruction from a methodological point of view, as confirmed by simulation approaches (unpublished data). These observations may be true even for some groups placed as early branches with high BPs, such as the microsporidians, diplomonads, and trichomonads.

EF-1a as a Phylogenetic Marker

The eukaryotic EF-1a-based phylogeny exhibits important divergences from that based on rRNA sequences, some of which have also been shown by other phylogenetic markers and biological data. A clear example is the monophyly of mycetozoans and their inclusion within the crown (Baldauf and Doolittle 1997). However, several problems make the reliability of this marker questionable for the intermediate and basal regions of the eukaryotic tree, where most protist species are located.

As shown, this problem arose from two main factors: high mutational saturation and the long-branch attraction artifact. In fact, mutational saturation erases most of the phylogenetic signal, thereby allowing the long-branch attraction artifact to be the main factor in determining the affinities of the sequences during phylogenetic analysis. In addition, the outgroup used to build EF-1a-based eukaryotic phylogenies presents an interesting problem due to its sequence variability. Archaeal EF-1a sequences exhibit an extreme degree of variability, with 276 variable positions out of 387 aligned ones, the highest number noted among the groups analyzed (table 3). The sole well-conserved positions of the archaeal EF-1a correspond mainly to constant positions in eukaryotic EF-1a (fig. 4). These positions are scarcely informative for phylogenetic inference, since they do not allow differentiation among groups or they represent nonpolarizable characters. This may also increase the long-branch attraction phenomenon, since those highly variable eukaryotic sequences will be strongly attracted by the highly variable outgroup, as proven for other markers (Lockhart et al. 1996). This seems to especially affect the branches in the basal and middle regions of the tree, which appear more sensitive to sampling differences in the outgroup (Roger et al. 1999). In addition, the position-by-position variability analysis has shown that the large numbers of EF-1a positions free to vary, or covarions, are dissimilar among the different ciliate groups, as seen for the divergent *Euplotes* sequences, in which 32 of the 157 variable positions are constant within the rest of ciliates (fig. 4). This suggests the existence of a high degree of heterogeneity in the ciliate covarion composition, not only as compared with other eukaryotic groups, but also within this phylum. A complex set of factors may be related to this heterogeneity, such as the unequal rates of other elements of the ciliate translation machinery, especially the rRNAs, imposing different coevolutionary constraints on the EF-1a genes. This heterogeneous covarion composition has important consequences, since correction models such as the gamma distribution (Yang 1996), valid for sequences with homogeneous among-site variation, cannot be successfully applied to sequences with such heterogeneous covarion composition. In fact, our analyses of the EF-1a data set using this correction method did not overcome the problems discussed here, yielding results almost identical to those of the classical ML analysis (data not shown).

In contrast, the most apical region of the tree exhibits more consistency, probably owing to the fact that the level of mutational saturation is much lower there than in the rest of the tree. Therefore, the usefulness of EF-1a for the inference of the eukaryotic phylogeny cannot be fully rejected, since it seems to be a suitable marker for analyzing some phyla in the crown, as demonstrated by the phylogenetic analysis of the Mycetozoa, misplaced by some other markers (Baldauf and Doolittle
Another example may be the interesting grouping of Euglenozoa and angiosperms into a well-supported terminal crown assemblage (fig. 3), in contradiction to the rRNA-based phylogeny. Several biological data agree with this observation, although some others clearly differentiate the two groups (Cavalier-Smith 1987). Therefore, this putative assemblage should be cautiously regarded until the evaluation of additional phylogenetic markers leads to a definitive answer.

The comparison of different phylogenies suggests that no presently known marker avoids all phylogenetic reconstruction artifacts. All of them yield more-or-less contradictory results and display trees with presumably problematic regions (e.g., asymmetric bases). Most likely, a reasonable picture of eukaryotic evolution will be obtained only in the form of a composite that integrates the reliable nodes from each individual phylogeny. In this regard, EF-1α may be useful in unraveling the instability of several clades, while supporting the robustness of others.

**Acknowledgments**

We acknowledge the constructive criticisms of Jacqueline Laurent, Philippe Lopez, Purificación López-García, Miklós Müller, and Andrew Roger. We would like to thank Masumi Nakazawa for advice during the initial phase of this work. We are grateful to Anne Baroin-Tourancheau for providing *Spathidium* sp., *Kenophoros* sp. cells and to Ghislaine Fryd and Francine Iftode for the identification and growth of ciliates. We thank Philippe Lopez for his computer programs, Karine Budin for advice on figures, and Nöelle Narroad for technical help. This work was supported by a grant from the Réseau National de Biosystématique. D.M. is a post-doctoral fellow of the Spanish Ministerio de Educación y Cultura.

**LITERATURE CITED**


BERGEMANN, J., V. FLORIAN, T. KREMSER, and A. KLEIN. 1996. Two different macronuclear EF-1α-encoding genes of the ciliate *Euplotes crassus* are very dissimilar in their sequencs, copy numbers and transcriptional activities. Gene 168:109–112.


Manolo Gouy, reviewing editor

Accepted October 22, 1998