The EMBO Journal vol.14 no.13 pp.3262–3267, 1995

Genetic code deviations in the ciliates: evidence for multiple and independent events

Anne Baroin Tourancheau[1], Nora Tsao[2], Lawrence A.Klobutcher[3], Ronald E.Pearlman[2] and André Adoutte

Laboratoire de Biologie Cellulaire 4 (URA 1134), Bâtiment 444, Université Paris-Sud, 91405 Orsay Cedex, France, [2]Department of Biology, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada and [3]Department of Biochemistry, University of Connecticut Health Center, Farmington, CT 06030, USA.

[1]Corresponding author

In several species of ciliates, the universal stop codons UAA and UAG are translated into glutamine, while in the euplotids, the glutamine codon usage is normal, but UGA appears to be translated as cysteine. Because the emerging position of this monophyletic group in the eukaryotic lineage is relatively late, this deviant genetic code represents a derived state of the universal code. The question is therefore raised as to how these changes arose within the evolutionary pathways of the phylum. Here, we have investigated the presence of stop codons in α tubulin and/or phosphoglycerate kinase gene coding sequences from diverse species of ciliates scattered over the phylogenetic tree constructed from 28S rRNA sequences. In our data set, when deviations occur they correspond to in frame UAA and UAG coding for glutamine. By combining these new data with those previously reported, we show that (i) utilization of UAA and UAG codons occurs to different extents between, but also within, the different classes of ciliates and (ii) the resulting phylogenetic pattern of deviations from the universal code cannot be accounted for by a scenario involving a single transition to the unusual code. Thus, contrary to expectations, deviations from the universal genetic code have arisen independently several times within the phylum.

Key words: ciliates/evolution/genetic code/ribosomal RNA/ UAA and UAG reassignment.

Introduction

A remarkable feature of the ciliates is their genetic code which is one of the few deviating from the universal code. In several species (*Paramecium*, *Tetrahymena*, *Stylonychia* and *Oxytricha*), the universal termination codons UAA and UAG are translated into glutamine and UGA is the only termination codon (see Caron, 1990; Prescott, 1994 and references therein). Evidence was obtained that this deviant code is derived and not the relic of a primordial genetic code. The key argument is that ciliates emerge as a late branching group in the general tree of eukaryotes preceeded and followed by groups with a normal code as first stressed by Baroin *et al.* (1988). In addition, the basis for this non-universal genetic code was elucidated in *Tetrahymena* and is compatible with its late emergence. In *Tetrahymena*, in addition to the normal glutamine $tRNA_{UmUG}^{Gln}$ which reads CAA and CAG codons, two unusual glutamine-specific tRNAs were isolated ($tRNA_{UUA}^{Gln}$ and $tRNA_{CUA}^{Gln}$). From the similarity of the three nucleotide sequences it was concluded that the two new glutamine tRNAs which recognize UAA and UAG arose from the normal one by gene duplication and divergence (Hanyu *et al.*, 1986; Jukes *et al.*, 1987). Deviant glutamine codons are not, however, found in all ciliates. In several representatives of the ciliate genus *Euplotes*, glutamine has been shown to be encoded by only the universal CAA and CAG codons. In this genus, UAA and UAG are used as stop codons and UGA appears to be translated as cysteine in *Euplotes octocarinatus* and *Euplotes crassus* (Meyer *et al.*, 1991, 1992; Jahn *et al.*, 1993). Similarly, in the heterotrich *Blepharisma japonicum*, it was recently shown that UAA is used as a stop codon (Liang and Heckmann, 1993). It is not known whether UAG and UGA are used as sense codons in this species. Thus, the genetic code has evolved even within the phylum, and this raises the question of how many times changes in the UAA/UAG/UGA codon usage have occurred in the evolutionary history of the group.

We have approached this question by superimposing the pattern of deviant codon utilization in as many species as possible over an independently established molecular phylogeny of these species. Indeed, fairly extensive and congruent molecular phylogenies based on either 18S (Greenwood *et al.*, 1991; Leipe *et al.*, 1994) or 28S (Baroin Tourancheau *et al.*, 1992) ribosomal RNA have recently been obtained. In the present limited sample of species where codon usage is known, *Blepharisma* represents the earliest diverged lineage. The heterotrichs indeed form a monophyletic lineage emerging at the base of the phylum (Baroin Tourancheau *et al.*, 1992). The apparently non-deviated code of *Blepharisma* would thus correspond to the ancestral state of the genetic code in the phylum as suggested by Liang and Heckmann (1993). However, an early origin of the deviant code followed by reversion in the *Blepharisma* and *Euplotes* lineages has also been proposed (Prescott, 1994). This alternative may be resolved by extending the codon usage data base to species belonging to all the major groups of the ciliates, including representatives from the most ancient ones.

In this study we have undertaken an analysis of the evolution of the genetic code by comparing two experimental data sets. First, we have sequenced α tubulin and/or phosphoglycerate kinase (PGK) genes from eight and five species respectively, which allows us to extend the genetic code data to six classes of the phylum (Hypotrichea, Oligohymenophorea, Litostomatea, Nassophorea,

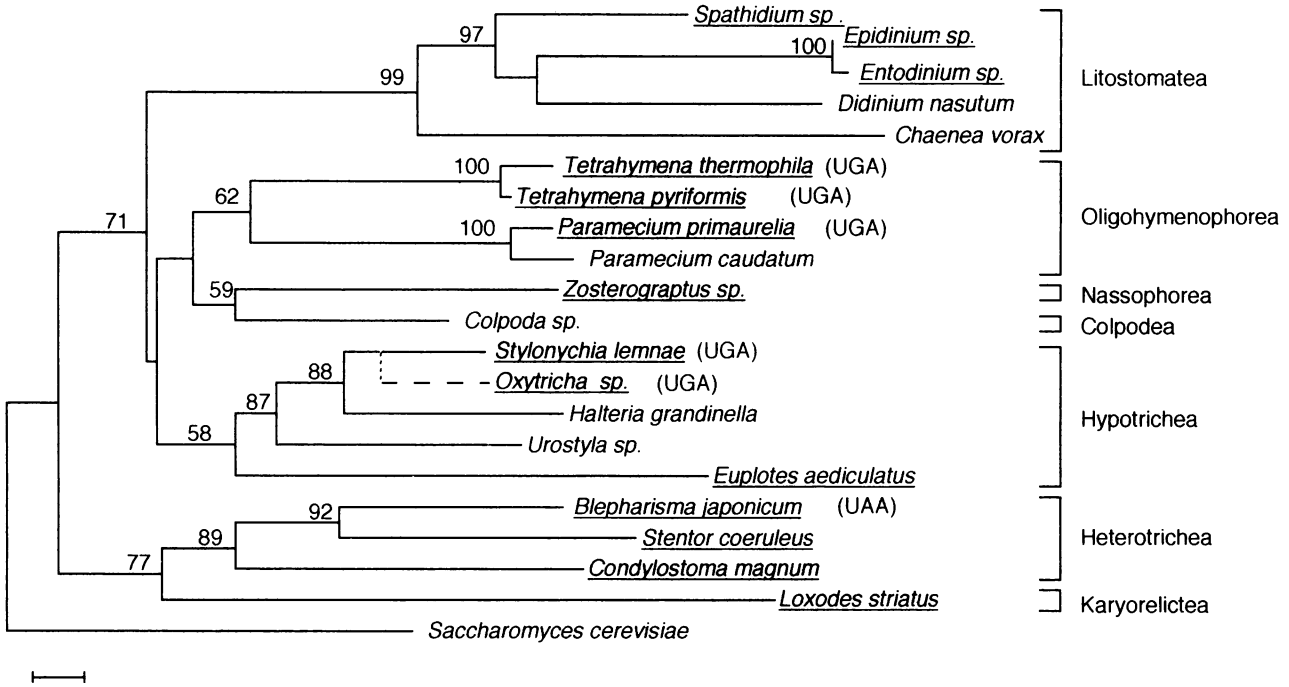

Fig. 1. Phylogenetic pattern of the genetic code in ciliates. Phylogenetic tree of the ciliates generated by the Neighbor-Joining method. The analysis is restricted to the 5′ end of the 28S rRNA molecule (see Materials and methods for details). Among the 250 unambiguously aligned sites, 136 are variable. Representatives from seven classes are included in the tree. *Saccharomyces cerevisiae* is taken as an outgroup. The 28S rRNA sequence of *Oxytricha sp.* is not available. Its position (marked by a broken line) is indicated in reference to its position in the 18S rRNA tree (Greenwood *et al.*, 1991; Leipe *et al.*, 1994). All substitutions are taken into account in the matrix data. The few gaps postulated in the selected domains that correspond to a single nucleotide were treated as a fifth base. Bootstrap values (>55%) that corroborate topological elements in the figure are indicated above the branches. Species underlined in colour are species in which the glutamine residues are encoded by universal codons and by the additional codons UAA and/or UAG. Black underlined species use, as so far determined, the universal glutamine codons. Termination codons when known are indicated. Scale bar, 1% of nucleotide substitutions.

Heterotrichea and Karyorelictea). These data show that deviations are extensively present in the phylum. Deviation from the universal code is by far the most common for glutamine, since deviant cysteine codons have only been observed in two species of the genus *Euplotes*. Second, we have superimposed the new and previous stop codon usage data on a phylogenetic tree constructed from partial sequences of 28S rRNA molecules. In the resulting phylogenetic pattern of deviation from the universal genetic code, we do not find a simple split between one group of species with a deviant code and one using the universal code, but rather a complete intermingling of the two types even within solid monophyletic groups. We are thus led to propose a polyphyletic origin of the code changes in the phylum.

Results

Ciliates are presently subdivided into eight classes (see Baroin Tourancheau *et al.*, 1992) but information on the ciliate codon usage has been restricted to representatives of three hypotrichs (*Euplotes*, *Stylonychia* and *Oxytricha*) and of two oligohymenophorans (*Paramecium* and *Tetrahymena*). In our data set, we have extended the genetic code analysis to species from four other classes: the karyorelictid *Loxodes striatus*, the two heterotrichs *Condylostoma magnum* and *Stentor coeruleus*, the nassophorean *Zosterographus sp.* and the three litostomes *Spathidium sp.*, *Epidinium sp.* and *Entodinium sp.*. The present phylogenetic tree of these classes is recapitulated in Figure 1. Current 18S and 28S ribosomal RNA phylogenies of the ciliates provide a very congruent pattern of diversification consisting of several well-resolved major branches which correspond to most of the morphologically defined classes (Greenwood *et al.*, 1991; Baroin Tourancheau *et al.*, 1992; Leipe *et al.*, 1994). The solid topological features of the rRNA trees, some of which appear in Figure 1, are the following: the bootstrap values (88 and 77%, respectively) strongly support the monophyly of the heterotrichs and their association with the karyorelictids. This lineage emerges as a sister group of all the other ciliates. The litostomes, the hypotrichs and the oligohymenophorans form three monophyletic groups supported in 99, 58 and 61% respectively, of the replicates. The monophyly of the oligohymenophoran, colpodean, nassophorean assemblage does not appear robust (28%). However, when using complete 18S rRNA sequences, the association of colpoda with the oligohymenophorans is supported with statistically significant bootstrap values (Leipe *et al.*, 1994). The litostomes, hypotrichs and oligohymenophorans + colpodeans + nassophoreans separate in an unresolved branching order (18%). The lack of resolution of this node is a constant feature of the ribosomal RNA phylogenies.

In the current analysis, all the new α tubulin sequences were obtained by PCR amplification. The PGK sequences were obtained either by PCR amplification or by screening genomic or cDNA libraries (see Materials and methods). For both proteins, the PCR products encode peptide sequences of ~400 amino acids containing ~12 glutamine residues. We took advantage of the fact that α tubulins

Table I. Glutamine codon usage in α tubulin and phosphoglycerate kinase genes

	CAA	CAG	UAA	UAG	STOP
Hypotrichea					
Euplotes aediculatus (α)	9	2	0	0	?
Euplotes vannus (α)	8	5	0	0	UAA
Euplotes octocarinatus (α)	12	1	0	0	UAA
Euplotes crassus (PGK)	7	1	0	0	UAG
Oxytricha granulifera (α)	11	0	3	0	UGA
Oxytricha nova (PGK)	6	3	1	3	UGA
Stylonychia lemnae (α)	11	0	1	0	UGA
Oligohymenophorea					
Tetrahymena pyriformis (α)	12	1	0	0	UGA
Tetrahymena thermophila (α)	7	0	5	1	UGA
Paramecium tetraurelia (α)	2	1	10	0	UGA
Tetrahymena pyriformis (PGK)	7	0	1	3	(UGA)
Tetrahymena thermophila (PGK)	4	0	6	0	UGA
Paramecium primaurelia (PGK)	0	0	10	2	(UGA)
Glaucoma chattoni (PGK)	2	1	1	2	(UGA)
Nassophorea					
Zosterograptus sp. (α)	6	0	6	0	?
Heterotrichea					
Condylostoma magnum (α)	4	3	5	0	?
Stentor coeruleus (α)	12	0	0	0	?
Condylostoma magnum (PGK)	4	6	4	1	(UGA)
Karyorelictea					
Loxodes striatus (α)	6	0	4	1	(UGA)
Litostomatea					
Spathidium sp. (α)	12	0	0	0	?
Epidiniumsp. (α)	12	0	0	0	?
Entodinium sp. (α)	11	0	0	0	?

The data from underlined species have been obtained in this study. Other data are from the literature or EMBL data library releases. The codon usage in α tubulin genes has been examined on a total of 380 amino acids (underlined data) and 450 amino acids (EMBL data), respectively. Of the seven ciliate classes molecularly defined (Baroin Tourancheau *et al.*, 1992), six are represented and indicated in bold characters. The termination codons are indicated when directly determined and in parentheses when inferred.

are highly conserved among species, so that in ciliates the alignment of the sequences at both the nucleotide and amino acid level is straightforward. As a result, the finding of any deviant codons can be used to infer the amino acid that is encoded with a high degree of certainty. An examination of all the α tubulin genes sequenced indicates that all in frame UAA and UAG triplets observed exist at positions which normally encode glutamine. PGK, a glycolytic enzyme, is a less conserved protein than α tubulin so that assignment of an in frame termination codon to a given amino acid may be ambiguous. However, an almost completely conserved glutamine residue among eukaryotic PGKs which occurs in the highly conserved sequence KIQLI (Vohra *et al.*, 1992) is encoded by UAA in both *Tetrahymena thermophila* and *Paramecium primaurelia* and by universal codons in the other ciliates. Also, examination of codon usage for other moderately, but not completely conserved, glutamine residues is consistent with the translation of UAA and UAG into glutamine for the PGKs in the other species analyzed. In frame UGA codons were not observed in any of the tubulin and PGK sequences.

Table I shows the codon usage of glutamine as deduced from our new sequences. It appears from these data that the α tubulin and PGK genes have a similar glutamine codon usage. In species or genera in which the two genes are sequenced, the non-standard codons are used at an approximately equal frequency. One exception is *Tetrahymena pyriformis* in which no deviation is present in the α tubulin gene, but deviations are found in the PGK gene, the β tubulin gene (Barahona *et al.*, 1988) and the actin gene (Hirono *et al.*, 1987). Comparison of the codon usage between the different classes indicates a widely distributed usage of deviant glutamine codons (predominantly UAA) over the entire phylum. In hypotrichs and oligohymenophorans, our results confirm previously reported data from other genes. Non-standard glutamine codons are not used in the four species of *Euplotes*, whereas in *Oxytricha* and in *Stylonychia lemnae*, UAA encodes glutamine as does UAG in *Oxytricha nova*. In oligohymenophorans, non-standard codons are more frequently used. The PGK sequences show deviations in all representatives of this class with a large bias in *Paramecium* towards UAA, as earlier reported in surface antigen, calmodulin and tubulin genes (Prat *et al.*, 1986; Preer *et al.*, 1987; Kink *et al.*, 1990; Dupuis, 1992). In the nassophorean, in one of the two heterotrichs (*C.magnum*) and in the karyorelictid, UAA, and to a lesser extent UAG, are consistently used to encode glutamine. However, in the heterotrich *S.coeruleus* and in the three

litostomes, species for which only the α tubulin sequences are available, no deviation is found. For these 'apparently orthodox' species, determination of the translation termination signal, as well as additional sequence data, will be necessary to confirm a universal pattern of glutamine codon usage. From this investigation, it is clear that the genetic code deviates within the six classes of ciliates except possibly for Litostomatea. What seems to vary is the extent to which the deviant code is used between different species within each class. It ranges, in hypotrichs for example, from normal usage (*Euplotes*) to moderate abnormal usage (*Stylonychia*) or, in heterotrichs, from normal, or apparently normal (*Stentor*), to highly deviant (*Condylostoma*).

In Figure 1 we have superimposed these data on the 28S rRNA based phylogeny of the phylum. Two major observations emerge from this analysis. First, the resulting phylogenetic pattern of the glutamine codon usage does not show a clear-cut separation between 'deviant' species and 'orthodox' or 'apparently orthodox' ones, but rather a mixing of both kinds of species even within a monophyletic lineage. This latter point is particularly well illustrated within the heterotrichs. As already mentioned, it has been shown that *B.japonicum*, which is a close relative to *S.coeruleus*, uses UAA as a termination codon (Liang and Heckmann, 1993). In *S.coeruleus*, UAA is not yet allocated but has not been found in frame. In contrast, in *C.magnum*, a more distant heterotrich relative, UAA is used as a sense codon. Thus, within a solid monophyletic group, two types of codon usage are found. A similar situation was already described in the hypotrichs. In contrast to what is observed in the closely related species *Stylonychia* and *Oxytricha* which contain in frame UAAs and UAGs, the codon UAA and also UAG are termination codons for the deeply diverged *Euplotes* species, while UGA was recently shown to encode cysteine in the pheromone gene of *E.octocarinatus* (Meyer *et al.*, 1991, 1992). Jahn *et al.* (1993) have also presented evidence for UGA encoding cysteine in the Tec transposable elements of *E.crassus*, whereas in the PGK gene (this work), no such deviation is observed. This type of deviation is also absent from all the tubulin genes of *Euplotes* analyzed to date. We thus note that in the euplotids, the tubulin and PGK genes do not seem to be useful indicators of the propensity of these species to use a non-universal code.

Second, the present work establishes that the deviant code is used in species belonging to the earliest emerging branch of the phylum and especially in the karyorelictids which are thought to derive from the most ancestral stock of ciliates. The combination of these two points strongly suggests that evolution of codon assignment has occurred independently several times within the phylum of ciliates. Indeed, there is no parsimonious way to account for the present distribution by, for example, a single transition from a normal to a deviant code prior to the emergence of the litostome/hypotrich/oligohymenophoran radiation and followed by a reversion in the *Euplotes* lineage as we initially expected to find. Such a simple scenario, which postulated a normal code in the ancestral lineage of the phylum and which was strengthened by the finding of UAA as the termination codon in *B.japonicum*, is ruled out by our present data. Another possibility is to postulate a transition from normal to deviant code prior to the emergence of the karyorelictid/heterotrich lineage followed by reversions to the normal code in diverse species (*Euplotes*, *Blepharisma*, possibly *Spathidium* etc.). This alternative requires numerous reassignments from glutamine to stop which is much more difficult to conceive. We therefore postulate multiple changes of the termination codons to glutamine codons occurring independently in distinct lineages as opposed to a single reassignment event. The variable degree with which these deviations are observed in the different species also fits with the notion of multiple independent reassignment events. The 'prociliate ancestor' would therefore have possessed a potential for evolution towards the development of a deviant code, and this potential would have been diversely realized among the different lineages.

Discussion

It is well documented that reassignment of codons has occurred in the course of evolution. Changes have been reported in many mitochondrial genomes and in the nuclear system of some organisms (see Osawa *et al.*, 1992). The only two other known deviant eukaryotic nuclear genomes are the asporogenic yeast *Candida cylindracea* and the green algae *Acetabularia* (Kawaguchi, 1989; Schneider, 1989). The data presented in this paper clarify two features of the unusual genetic code in ciliates: (i) the deviant code is not restricted to the most evolved species and (ii) different assignment patterns can be observed between related species which emphasizes the great flexibility of the reassignment process.

How might the current hypotheses on the evolution of the genetic code explain the multiple codon changes observed in the ciliates? Two major hypotheses have been presented. In the first, the codon capture hypothesis (Osawa and Jukes, 1989), mutational pressure towards an AT or GC rich genome is postulated to result in the complete disuse of a codon, followed by the loss of the tRNA responsible for translation of the codon. It is then possible for the codon to reappear in conjunction with a duplicated and mutated tRNA such that codon reassignment results. In the case of the glutamine codons (stop codon capture), Jukes *et al.* (1991) proposed that the polypeptide chain release factors would have become specific for UGA, so that UAA and UAG became unrecognizable and disappeared from the genome. Under AT pressure, some of the CAA and CAG glutamine codons mutated to UAA and UAG, while the gene for glutamine tRNA duplicated and mutated giving rise to the two new tRNAs with their respective UUA and CUA anticodons pairing with UAA and UAG. In the euplotid lineage, a slightly different course of events would have occurred with the polypeptide release factors becoming specific for UAA and UAG, therefore releasing UGA and allowing its reassignment to cysteine. Based on this model, the phylogenetic termination codon usage pattern in Figure 1 would reflect the multiple changes in the specificity of the polypeptide release factors leading to multiple and independent reassignments of the codons specifying the termination of protein synthesis. Ciliates are known to have AT rich genomes and a drift in the base composition of the genome could indeed have contributed to these reassignments. In this case, we expect to find a correlation between an

increase in AT pressure and an increase in the use of UAA termination codons as sense codons. In hypotrichs, in which deviant codons are not frequently used, it seems that macronuclear non-coding DNA is not as A + T rich as in *Paramecium* and *Tetrahymena* (Prescott, 1994). The following arguments do not support a strong AT driving force. First, the frequency of utilization of UAA in *Paramecium* and in *Tetrahymena* is very different although these organisms have similar very rich A + T genomes. Second, in absence of data on the genomic base composition of the other groups of ciliates, we have calculated the G + C content in the divergent domains of the 28S ribosomal molecules (which are less constrained than the conserved domains of the molecules) for each species, and also the third position G + C content of the tubulin and PGK genes (data not shown). From these estimations, we cannot find a correlation between the extent to which a deviant code is used and an AT mutational bias.

In a different scenario, Anderson and Kurland (1990) suggest that evolution of the translation system could take place gradually through mutation of the relevant tRNA species. The second hypothesis, tRNA-mediated codon reassignment (Schultz and Yarus, 1994), supports this view and postulates that the crucial intermediate in the evolution of the code is the mutation in tRNA that allows the tRNA to simultaneously recognize cognate and near-cognate codons. In this model, reassignments of codons are proposed to occur through transitional translationally ambiguous periods during which codons have two different meanings. Stop codons do not have cognate tRNAs and their reading by near-cognate tRNA is quite possible since natural suppressor tRNAs have been shown to exist in many organisms. In the case of glutamine codons, UAA and UAG may have temporarily simultaneously encoded stop and glutamine before their definitive assignment to glutamine. In support of this model, Schultz and Yarus (1994) have noted that all three isoacceptors of $tRNA^{Gln}$ in *T.thermophila* are predicted to allow near-cognate recognition of UAA and UAG codons. These $tRNA^{Gln}$ are presumed to have diverged from a mutated ancestor which could translate the near cognate UAA/UAG codons in addition to the cognate CAA/CAG codons. The authors therefore suggest that these sequence changes potentiated the observed changes in the genetic code of *T.thermophila*. If these changes occurred in the $tRNA^{Gln}$ gene(s) of an ancestral ciliate utilizing the standard genetic code, they may well have set up a situation whereby multiple independent changes in the genetic code could easily occur as we observe and report here. It will be of interest to examine $tRNA^{Gln}$ genes in a number of ciliate groups to test this hypothesis.

Whatever the mechanisms that have led to the abnormal codon usage in ciliates, one may wonder why these organisms have shown such a propensity for developing these deviations in several independent lineages. Could there be a selective advantage in ciliates favoring such an evolution? For example, in the case of the mitochondrial code, a strong point could be made for a selective advantage of codon reassignment since it allows substantial reduction in genome size (Anderson and Kurland, 1990). Ciliates are very active phagotrophs. All the species in the phylum, except highly derived forms, possess a specialized organelle known as the oral apparatus for active digestion of both bacteria and other eukaryotes. This has led Ninio (1986) to speculate that a deviant code was favored in ciliates because it constitutes a mechanism preventing the expression of foreign DNA. Similarly, Prescott (1994) has noted that all ciliates studied to date are devoid of virus although bacterial symbionts are very common. These ideas may, however, be over simplifications since many ciliates contain transposable elements, at least some of which appear to be ancient. The extremely wide distribution of hosts for these elements suggest horizontal transfer, and that they entered the ciliate lineage utilizing the universal code (Doak *et al.*, 1993; Jahn *et al.*, 1993). Codon reassignment has, however, occurred in these elements as in other ciliate genes during evolution as evidenced by the transposase gene in Tec elements of *E.crassus* using UGA to encode cysteine (Jahn *et al.*, 1993). Alternatively, non-standard codon usage may be selectively neutral and may have been fixed through a process which is unique to ciliates. It is tempting to suggest that the occurrence of nuclear dimorphism, i.e. the existence of a 'silent' germline micronucleus and of a highly polyploid transcriptionally active macronucleus within the same cell (a property which is both common to all ciliates and unique to this phylum), has facilitated in one of several possible ways, the drift towards abnormal codon usage. Whatever the final explanation, ciliates as a whole provide a remarkable material for investigating evolution of the genetic code and may in fact allow us to capture some of the intermediate stages in this evolution.

Materials and methods

Sequencing of α tubulin genes

α tubulin genes were sequenced by cloning PCR amplification products synthesized from genomic DNA isolated from frozen or ethanol fixed cells using 17 nuclotide degenerate primers. The sense and antisense primers commercially synthesized (Bioprobe, France) are based on the conserved CLEHGI and MEEGEF amino acid sequences. Amplifications were performed using 50 ng DNA in 50 µl of 10 mM Tris–HCl pH 8.3, 1.5 mM $MgCl_2$, 200 µM deoxyribonucleoside triphosphates, 1 µM each primer and topped with mineral oil. After a 3 min denaturation step at 94°C in a programmable thermalcycler (Crocodile, Appligene, France), 2 U of *Taq* polymerase (Bioprobe, France) was added and 35 cycles were carried out as follows: 1 min at 94°C, 1 min at 55°C and 1.5 min at 72°C. After a last elongation step at 72°C for 5 min, the 1.1 kb PCR products were purified using Jet pure (Bioprobe, France) and cloned into either the *Eco*RV site of the plasmid vector pBS (Stratagene, La Jolla, CA) or the *Sma*I site of the plasmid vector pUC (Pharmacia, France). Dideoxy double strand sequencing was performed using T7 DNA polymerase (Amersham, France) and standard procedures. For each species, the amino acid sequence was deduced from the sequence of a single clone. The amino acid sequences were aligned with the following α tubulin sequences taken from the EMBL data bank (with their respective accession numbers): *Euplotes vannus*, Z11769; *E.octocarinatus*, X69466; *S.lemnae*, X01746; *Oxytricha granulifera*, Z11763; *T.pyriformis*, X12767; *T.thermophila*, M86723; The sequence of *Paramecium tetraurelia* is taken from Dupuis (1992).

Sequencing of the PGK genes

The single copy PGK gene from *E.crassus* and *Oxytricha nova* was isolated by screening bacteriophage macronuclear libraries (Baird and Klobutcher, 1988; Klobutcher *et al.*, 1984) with a *T.thermophila* 900 bp *Eco*RI cDNA probe (Vohra *et al.*, 1992) at moderate stringency. The PGK genes from the other species were PCR amplified from DNA from frozen or ethanol fixed cells using degenerate primers (Dalton Chemical Co.) based on highly conserved regions of the protein (Vohra *et al.*, 1992). The sense 24 nucleotide primer is based on the amino acid sequence VDFNVP and the antisense 22 nucleotide primer on the sequence TGGGAS. Amplifications were performed using 100 ng DNA

in 25 µl of 10 mM Tris–HCL pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1 mg/ml gelatin, 200 µM deoxyribonucleoside triphosphates, 2 µM each primer and topped with mineral oil. A 'touchdown' procedure (Don *et al.*, 1991) was used with a Coy thermal cycler. Denaturation was initiated at 94°C for 4 min and 1 U of AmpliTaq DNA polymerase (Perkin-Elmer/Cetus) was added at 2 min. Thirty five cycles were carried out as follows: 94°C for 30 s, 64°C for 2 min, 72°C for 1.5 min for 1 cycle; 94°C for 30 s, 62°C for 2 min, 72°C for 1.5 min for 2 cycles; 94°C for 30 s, 60°C for 2 min, 72°C for 1.5 min for 2 cycles; 94°C for 30 s, 58°C for 2 min, 72°C for 1.5 min for 30 cycles; and 72°C for 10 min for 1 cycle. The ~1.1 kb amplification products were either cloned into the *Sma*I site or the *Eco*RI + *Bam*HI sites of the vectors pUC or pEMBL. Dideoxy sequencing of double-stranded DNA was done using the Sequenase Version 2 kit (USB) and 35SdATP (Dupont NEN) according to supplier's directions. Amino acid sequences were deduced from the sequence of a single clone. The ciliate PGK sequences were aligned using Clustal (Higgins and Sharp, 1989).

Sequencing of the 28S rRNA genes

The 5′ terminal domain of the 28S rDNA was amplified in the three litostomes (*Spathidium sp.*, *Epidinium sp.* and *Entodinium sp.*), in the nassophorean (*Zosterographus sp.*), in the colpodean (*Colpoda sp.*) and in the heterotrich (*C.magnum*), using primers which correspond to sequences at nucleotides 25–44 and 1126–1144 in *Mus musculus*. The amplification, cloning and sequencing procedures were identical to those described for the α tubulin genes except that the PCR cycles were performed as follows: 1 min at 94°C, 1 min at 50°C and 1 min at 72°C. The sequences from the other ciliates except *Tetrahymena* were obtained through the direct sequencing of the RNA molecules as previously described (Baroin Tourancheau *et al.*, 1992).

Construction of the phylogenetic tree

The alignment of the 28S rDNA sequences was performed 'by eye'. They were aligned with homologous 28S rRNA sequences previously obtained from the following ciliates and available from GenBank (with their respective accession numbers): *L.striatus*, M98368; *S.coeruleus*, M19220; *B.japonicum*, M19221; *Chaenea vorax*, M98365; *Didinium nasutum*, M98362; *Euplotes aediculatus*, M98383; *S.lemnae*, M98381; *Halteria grandinella*, M98371; *Urostyla sp.* M98379; EMBL accession numbers are X54004 for *T.pyriformis*, X54512 for *T.thermophila* and J01355 for *Saccharomyces cerevisiae* taken as a distantly related organism. Parts of the sequenced domains are highly variable and can be aligned only between closely related species. Only unambiguously aligned regions were retained for phylogenetic analyses. Editing of the sequences, storage, choice of the domains analyzed, formatting for the distance–matrix tree building program and printing of the tree were carried out using the MUST package (Philippe, 1993). The phylogenetic tree is derived from distance matrix using the Neighbor-Joining (NJ) procedure of Saitou and Nei (1987). The molecular distance for each pair of species was calculated from the frequency of the number of differences (a gap at a given position is considered as one difference). Evaluation of the statistical validity of the nodes was performed using NJBoot which applies the bootstrap procedure (1000 replicates) on the NJ method.

Acknowledgements

We thank Francine Iftode for the isolation and identification of ciliates and Eduardo Villalobo for participating in part of these experiments. This work was supported by the CNRS (France), the University of Paris-Sud (France) and by grants from the Medical Research Council of Canada to R.E.P. and from the National Institutes of Health (USA) to L.A.K.

References

Anderson,S.G.E. and Kurland,C.G. (1990) *Microbiol. Rev.*, **54**, 198–210.
Baird,S.E. and Klobutcher,L.A. (1988) *J. Protozool.*, **35**, 459–465.
Barahona,I., Soares,H., Cyrne,L., Penque,D., Denoulet,P. and Rodrigues-Pousada,C. (1988) *J. Mol. Biol.*, **202**, 365–382.
Baroin,A., Perasso,R., Qu,L.H., Brugerolle,G., Bachellerie,J.P. and Adoutte,A. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 3474–3478.
Baroin Tourancheau,A. Delgado,P., Perasso,R. and Adoutte,A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9764–9768.
Caron,F. (1990) *Experientia*, **46**, 1106–1117.
Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G. (1993) *Proc. Natl Acad. Sci. USA*, **91**, 942–946.
Don,R.H., Cox,P.T., Wainwright,B.J., Baker,K. and Mattick,J.S. (1991) *Nucleic Acids Res.*, **19**, 4008.
Dupuis,P. (1992) *PhD Thesis*, Structure organisation et expression des gènes de tubuline chez la paramecie. Université Paris-Sud.
Greenwood,S.J., Schlegel,M., Sogin,M.L. and Lynn,D.H. (1991) *J. Protozool.*, **38**, 1–6.
Hanyu,N., Kuchino,Y. Susumu,N. and Beier,H. (1986) *EMBO J.*, **5**, 1307–1311.
Higgins,D. and Sharp,P.M. (1989) *Comput. Appl. Biosci.*, **5**, 151–153.
Hirono,M. (1987) *J. Mol. Biol.*, **194**, 181–192.
Jahn,C.L., Doktor,S.Z., Frels,J.S., Jaraczewski,J.W. and Krikau,M.F. (1993) *Gene*, **133**, 71–78.
Jukes,T.H., Osawa,S. and Muto,A. (1987) *Nature*, **325**, 668.
Jukes,T.H., Bessho,Y., Ohama,T. and Osawa,S. (1991) *Nature*, **352**, 575.
Kawaguchi,Y., Honda,H., Taniguchi-Morimura,J. and Iwasaki,S. (1989) *Nature*, **341**, 164–166.
Kink,J.A., Maley,M.E., Preston,R.R., Ling,K.Y., Wallen-Friedman,M.A., Saimi,Y. and Kung,C. (1990) *Cell*, **62**, 165–174.
Klobutcher,L.A., Jahn,C.L. and Prescott,D.M. (1984) *Cell*, **36**, 1045–1055.
Leipe,D.D., Bernhard,D., Schlegel,K.M. and Sogin,M.L. (1994) *Eur. J. Protistol.*, **30**, 354–361.
Liang,A. and Heckmann,K. (1993) *Naturwissenschaften*, **80**, 225–226.
Meyer,F., Schmidt,H.J. and Heckmann,K. (1992) *Dev. Genet.*, **13**, 16–25.
Meyer,F., Schmidt,H.J., Plümper,E., Hasilik,A., Mersmann,G., Meyer, H.E. Engström,A. and Heckmann,K. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3758–3761.
Ninio,J. (1986) *Biochem. Syst. Ecol.*, **14**, 455–457.
Osawa,S. and Jukes,T.H. (1989) *J. Mol. Evol.*, **28**, 271–278.
Osawa,S., Jukes,T.H., Watanabe,K. and Muto,A. (1992) *Microbiol. Rev.*, **56**, 229–264.
Philippe,H. (1993) *Nucleic Acids Res.*, **21**, 5264–5272.
Prat,A., Katinka,M., Caron,F. and Meyer,E. (1986) *J. Mol. Biol.*, **189**, 47–60.
Preer,J.R., Preer,L.B., Rudman,B. and Barnett,A. (1987) *J. Protozool.*, **34**, 418–423.
Prescott,D.M. (1994) *Microbiol. Rev.*, **58**, 233–267.
Schneider,S.U., Leible,M.B. and Yang,X.P. (1989) *Mol. Gen. Genet.*, **218**, 445–452.
Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
Schultz,D.W. and Yarus,M. (1994) *J. Mol. Biol.*, **235**, 1377–1380.
Vohra,G.B., Golding,G.B., Tsao,N. and Pearlman,R.E. (1992) *J. Mol. Evol.*, **34**, 383–395.

Received on February 21, 1995; revised on April 6, 1995