The EMBO Journal vol.15 no.9 pp.2285–2290, 1996

A non-canonical genetic code in an early diverging eukaryotic lineage

Patrick J.Keeling[1] and W.Ford Doolittle

Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

[1]Corresponding author

The nearly invariant nature of the 'Universal Genetic Code' attests to its early establishment in evolution and to the difficulty of altering it now, since so many molecules are required for, and depend upon, faithful translation. Nevertheless, variations on the universal code are known in a handful of genomes. We have found one such variant in diplomonads, an early-diverging eukaryotic lineage. Genes for α-tubulin, β-tubulin and elongation factor 1 alpha (EF-1α) from two unclassified strains of Hexamitidae were found to contain TAA and TAG (TAR) triplets at positions suggesting a variant code in which TAR codes for glutamine. We found confirmation of this hypothesis by identifying genes encoding glutamine-tRNAs with CUA and UUA anticodons. The α-tubulin and EF-1α genes from two other diplomonads, *Spironucleus muris* and *Hexamita inflata*, were also sequenced and shown to contain no such non-canonical codons. However, tRNA genes with the anticodons UUA and CUA were found in *H.inflata*, suggesting that this diplomonad also uses these codons, albeit infrequently. The high GC content of these genomes and the presence of two isoaccepting tRNAs compound the difficulty of understanding how this variant code arose by strictly neutral means.

Keywords: codon capture/diplomonad/genetic code/tRNA

Introduction

Diplomonads are predominantly parasitic, amitochondrial protists which have consistently been shown through analyses of ultrastructural characteristics and molecular phylogeny to be among the first lineages of eukaryotic cells to have diverged from the main eukaryotic trunk (Cavalier-Smith, 1993; Leipe *et al.*, 1993). Diplomonads are members of the metamonads which, together with the microsporidia and archamoebae, belong to an assemblage known as the Archezoa. Archezoa are thought to have arisen before the acquisition of the mitochondria, and to have retained many primitive features of the first nucleated cells (Cavalier-Smith, 1993). Much of the molecular data and almost all of the protein-coding sequences known for diplomonads come from a single genus, *Giardia*. As part of an effort to obtain a wider representation of protein-coding sequences within this group, we sequenced the genes for α- and β-tubulin and elongation factor 1 alpha (EF-1α) from several diverse diplomonads.

Among these were two *Hexamita* strains (ATCC 50330 and 50380), blood-borne and muscle parasites of Pacific and Atlantic salmon, respectively, whose genes we found to contain numerous in-frame termination (TAA and TAG) codons. We showed these to be sense codons in these genomes by identifying cognate tRNA genes, and a survey of tRNA genes throughout several diplomonads revealed that another species, *Hexamita inflata*, also probably uses TAA and TAG (TAR) glutamine codons.

While almost all known genomes employ the same ancestral genetic code, variant codes have been identified in one bacterial genome, three eukaryotic nuclear lineages and in mitochondria. One theory which provides a plausible route for the evolution of such variants from the universal code is 'codon capture', developed predominantly by Osawa and Jukes (reviewed in Osawa *et al.*, 1992). This model has the important advantage that it avoids selectively disadvantaged transition stages through a series of neutral steps. First, either by mutation pressure or chance, certain codons disappear altogether from all genes in a genome. Once this occurs, the genes encoding the tRNAs or release factor previously required to read these missing codons are superfluous, and may be inactivated or lost. The codon is now 'unassigned', but may reappear in the genome if a new tRNA which can recognize it arises fortuitously (for instance if a duplicate of a functional tRNA gene acquires an anticodon mutation). Such a tRNA gene suppresses the lethal effects of chance or pressure-driven mutations which reintroduce the missing triplet, thus 'capturing' the codon and establishing a new code.

The particular variation of the genetic code observed here has previously been observed only in very AT-rich nuclear genomes, where it is thought to have been favoured by the same directional mutation pressure that biased the genome's composition (Osawa and Jukes, 1989). There is no evidence of such AT pressure on the GC-rich genomes of these diplomonads; however, even in the absence of directional mutation pressure, mutations converting glutamine to amber or ochre termination codons are expected to occur with a higher than average frequency because they involve only a single transition each. The lesion of these diplomonad genomes appears to be that there is no unique force required to change the genetic code, but that any mutation occurring at a sufficiently high frequency has the potential to motivate a codon capture event.

Results

Hexamita strains use a variant genetic code

Genes for α- and β-tubulin and EF-1α were amplified from *Hexamita* strain ATCC 50330, a newly characterized organism for which no molecular data are known. In

general, these new sequences closely resemble their homologues from *Giardia lamblia*, with the exception of α-tubulin which was previously uncharacterized in diplomonads. The unexpected feature of these *Hexamita* genes is the appearance of amber (TAG) and ochre (TAA) codons at several positions where glutamine is found conserved among diverse homologues. To extend this observation, we also isolated and sequenced the α-tubulin gene from another *Hexamita* strain (ATCC 50380), parasitic in Atlantic salmon. This second tubulin gene proved to be very similar to the first, differing at only three out of 1153 positions. Interestingly, two of these substitutions are synonymous transitions, while another interconverts glutamine and amber, also by a pyrimidine transition (Figure 1). If the two strains share the same variant code (as we suggest), this latter is also a synonymous substitution.

TAR codons also specify glutamine in *Acetabularia* and certain ciliates, where the conclusion that these triplets (in addition to CAG and CAA) code for glutamine has been confirmed by comparisons of gene and protein sequences (Schneider *et al.*, 1989), and most convincingly by the finding of $tRNA^{Gln}$ species with UUA and CUA anticodons in *Tetrahymena thermophila* (Hanyu *et al.*, 1986). We reasoned that compelling confirmation that the amino acid sequences shown here bespeak a similar variant code might be best obtained by using polymerase chain reaction (PCR) to search these *Hexamita* genomes for genes that encode novel tRNAs able to decode UAG and UAA. Such genes should generate PCR products of 83 bp, with the anticodon 13 bp from the terminus of the 5′ primer, and should be foldable into cloverleaf tRNA structures.

Indeed, products matching these criteria were readily obtained and sequencing indicated that they fell into three distinct groups: two of $tRNA^{Gln}$-like genes with the anticodons CTA and TTA, and one $tRNA^{Gly}$-like gene with the anticodon GCC. When aligned to a sample of all types of tRNAs from diverse organisms, the nearest relatives to the putative $tRNA^{Gln}$ fragments were glutamine tRNAs from other eukaryotes. The primary and predicted secondary structures show many other tRNA-like features, including the presence of an invariant U33 residue which is necessary to allow wobble pairing at position 34, a pyrimidine at position 32 and a purine at 37, as well as scattered conserved nucleotides and an anticodon stem, all spaced exactly as expected of a eukaryotic $tRNA^{Gln}$ (Figure 2).

The genetic code of other diplomonads

To see whether this curious trait is restricted to the closely related *Hexamita* strains, genes for α-tubulin and EF-1α were amplified and sequenced from *H.inflata* and *Spironucleus muris*. These are also shown in Figure 1, where it can be seen that no termination codons were observed in 21 glutamine codons of *H.inflata* or 16 glutamine codons of *S.muris*. However, the tRNA genes from these diplomonads are more revealing.

Even a slight decrease in the frequency of TAR use observed in the *Hexamita* strains (one out of 10 glutamine codons in ATCC 50330) could render them difficult to detect in protein coding sequences. In *H.inflata*, for instance, the frequency of TAR glutamine codons could

Fig. 1. Deduced amino acid sequences of EF-1α, β-tubulin and α-tubulin from *Hexamita* strains ATCC 50330, ATCC 50380, *S.muris* and *H.inflata*. Sequences are aligned to homologues from *G.lamblia* where possible, and in the case of α-tubulin to *Saccharomyces cerevisiae*. Termination codons are represented by a dot (•) and are also shown as triplets beneath the alignment. GenBank accession numbers for new sequences are U29440–2, U30664 and U37078–81. Abbreviations: *G.la*, *Giardia lamblia*; *S.ce*, *Saccharomyces cerevisiae*; *H.*30, ATCC 50330; *H.*80, ATCC 50380; *S.mu*, *Spironucleus muris*; *H.in. Hexamita inflata*.

be as high as one out of seven and it would still not be unlikely to observe none among the 21 glutamine codons encountered (based on a Poisson distribution of hits). tRNA genes were therefore amplified from *S.muris*, *H.inflata* and also *G.lamblia*, for which there is enough molecular data (including data on termination codons used at the ends

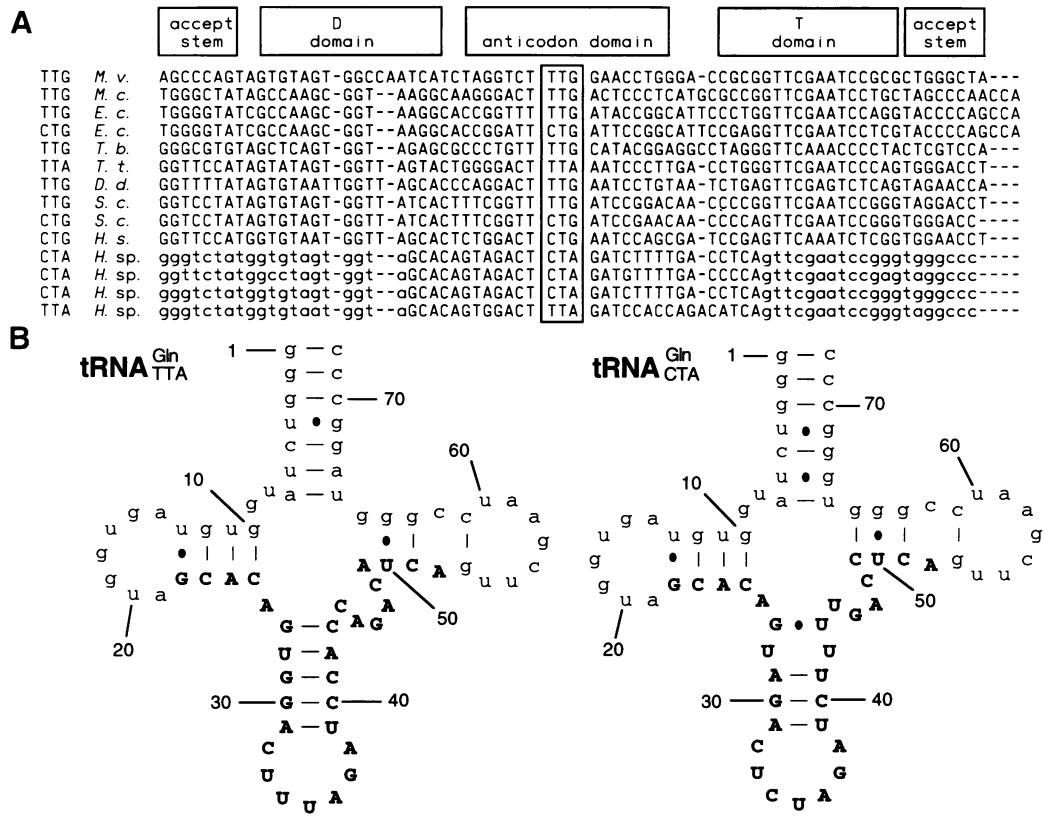

Fig. 2. Primary and secondary structure model of putative $tRNA^{Gln}$ species from *Hexamita*. (**A**) Nucleotide sequences from selected $tRNA^{Gln}$ genes aligned with amplification products from *Hexamita* 50330. Primer sequences are shown in lower case, domains indicated above the sequences and the anticodon distinguished by a box. (**B**) Proposed cloverleaf structures for $tRNA^{Gln}_{UUA}$ and $tRNA^{Gln}_{CUA}$ numbered according to the scheme of Sprinzl *et al.* (1989). Primer sequences are in lower case, amplified sequences are upper case and boldface. The primers have been included to give the structural context of the amplification product. *M. v.*, *Methanococcus vannielii*; *M. c.*, *Mycoplasma capricolum*; *E. c.*, *Escherichia coli*; *T. b.*, *Trypanosoma brucei*; *T. t.*, *Tetrahymena thermophila*; *D. d.*, *Dictyostelium discoideum*; *S. c.*, *Saccharomyces cerevisiae*; *H. s.*, *Homo sapiens*; *H.* sp., *Hexamita* strain ATCC 50330.

of open reading frames) to conclude that it uses the universal code.

The amplification products are aligned with those of *Hexamita* ATCC 50330 in Figure 3. *Giardia lamblia* yielded only a single $tRNA^{Gln}_{UUG}$ as well as a $tRNA^{Pro}_{UGG}$, but no non-canonical tRNAs. *Spironucleus muris* yielded only a single unambiguous product, corresponding to $tRNA^{Gln}_{CUG}$. The most interesting results were obtained from the free-living diplomonad, *H.inflata*. In this case, despite the fact that no termination codons were observed in the 783 codons comprising α-tubulin and EF-1α, genes for tRNAs that decode both TAA and TAG were readily obtained (in this case, as in *Hexamita* 50330, no canonical tRNAs were observed, but a $tRNA^{Met}_{CAU}$ was spuriously amplified).

Phylogenetic relationships between the diplomonads

Phylogenetic trees based on EF-1α amino acid sequences were inferred using maximum likelihood (Figure 4A), parsimony and distance (Figure 4B). The large data set analysed by parsimony and distance methods confirms the very early divergence of diplomonads, and argues very strongly that diplomonads are a monophyletic taxon. In addition, all methods gave the same topology for the diplomonads, although the statistical support for the relationship between *H.inflata* and *Hexamita* 50330 is very

Glutamine-Decoding tRNA Genes

*Hs*Q1	GCACAGTGGACT--TTA--GATCCACCAGACATCA	x2
*Hs*Q2	GCACAGTAGACT--CTA--GATCTTTT-GACCCCA	x3
*Hs*Q3	GCACAGTAGACT--CTA--GATCTTTT-GACCTCA	
*Hs*Q4	GCACAGTAGACT--CTA--GATGTTTT-GACCCCA	
*Hi*Q1	GCACACTGGACT--TTA--GATCCCGCAGACTTCG	
*Hi*Q2	CCACACTGGACT--CTA--GATCCCGCAGACTCCG	
*Hi*Q3	GCACACTGGACT--CTA--GATCCCGCAGACTCCG	
*Sm*Q1	GCATTTTTGATT--CTG--GTTCAAAC-GTCCCCG	
*Gl*Q1	TCACTTTCGGTT--TTG--ATCCGGACA-ACCCCG	x4

Non-Glutamine-Decoding tRNA Genes

*Hs*G1	GAATATGTGCTT--GCC--ATGCACGT-GGTCTGG	Gly
*Hi*M1	GGGCGCCAGGCT--CAT--AACCTGGAACTTGTTG	Met
*Gl*P1	TGATTCTCGCTT--TGG--GTGCGAGAGGTCCCGG	Pro

Fig. 3. Primary structure of putative $tRNA^{Gln}$ from four diplomonads. Sequences of all unique amplification products aligned against putative tRNA genes from ATCC 50330. In each case, the sequences may be folded into stem–loop structures resembling anticodon stems (except for *Hs*Q4 which only makes a potentially unstable stem, and may be a pseudogene) with the triplet 13 bp from the 5′ end in the anticodon position.

weak in parsimony and distance analyses (47 and 46%, respectively). Nevertheless, the support for this relationship is highly significant in maximum likelihood (estimated

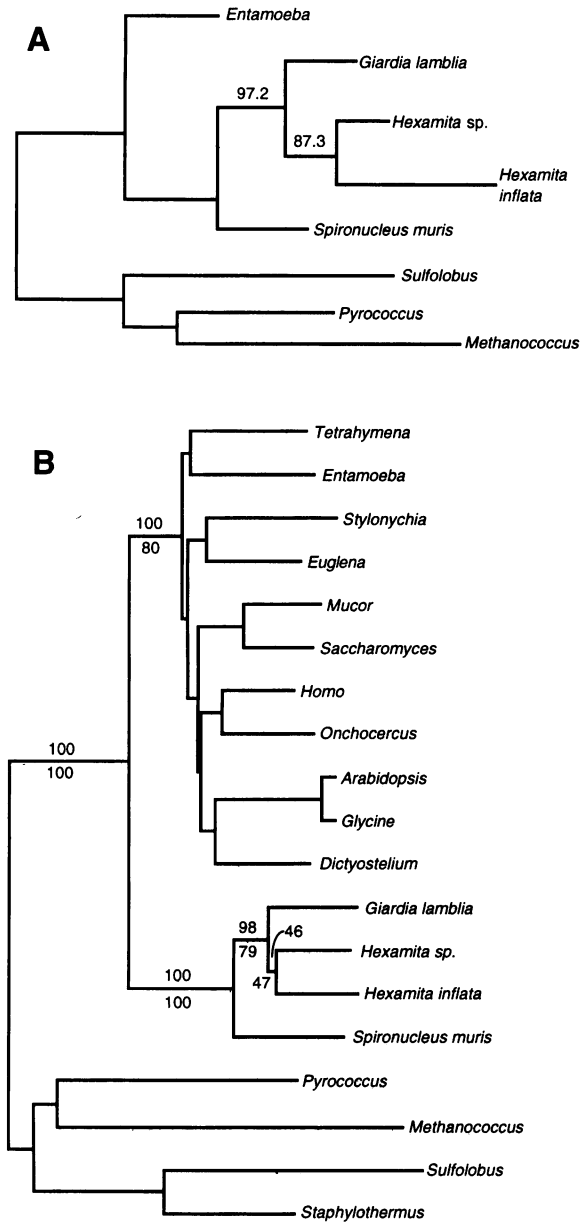

Fig. 4. Phylogenetic tree of selected EF-1α sequences. (**A**) Maximum likelihood topology of a restricted data set. Estimated bootstrap per cent is shown for all unconstrained nodes. (**B**) Distance topology shown with branch lengths. A nearly identical topology was also found for the most parsimonious tree, differing only in that *Entamoeba*, *Tetrahymena*, *Stylonychia*, and *Euglena* were multiply paraphyletic. Bootstrap per cents for nodes within and immediately surrounding the diplomonads are shown on the figure (distance above the node, parsimony below), others are excluded for clarity.

bootstrap of 87% and a 96% confidence that this topology is superior to any other), which has been shown to be more consistently correct in inferring relationships when rates are unequal between taxa or between sites within a taxon (Hasegawa and Fujiwara, 1993). In general, EF-1α phylogeny tends to support the conclusion that the two taxa with the non-canonical genetic code are themselves a clade. This has also now been seen in pylogenies based on GAPDH (Rozario *et al.*, 1996), but there must be more taxa included in both data sets before any firm conclusions can be drawn. In ciliates, it has become clear that the same variant code evolved in several groups independently (Tourancheau *et al.*, 1995), but this does not seem to be the case in these diplomonds.

Discussion

Changes to the genetic code in the nucleus are very rare. This case is only the fifth to be discovered and, interestingly, three of the others also involve TAA and TAG stop codons (both) changing to glutamine. The common involvement of termination codons in code alterations might be explained by their relatively low frequency, their functional redundancy and the fact that occasional failure to terminate translation of some proteins following loss of release factors should be less detrimental than the failure to complete translation of some proteins because of loss of tRNAs—factors which mitigate the effects of their loss. However, in eukaryotes, the specific and simultaneous capture of TAA and TAG by glutamine suggests that some further special relationships exist between these codons: no variant codes in which either TAR codon has been replaced by an amino acid other than glutamine have been described, and TAA and TAG seem always to be replaced together. We address these issues in turn.

Why always glutamine? The glutamine-encoding TAA and TAG in *Hexamita* genes presumably arose from CAA and CAG codons. Other organisms (ciliates and *Acetabularia*) where TAR codes for glutamine are very AT rich (as high as 76%; Schneider *et al.*, 1989; Prescott, 1994), and this has led to the suggestion (Osawa and Jukes, 1989; Osawa *et al.*, 1992) that the AT mutation pressure which biased the overall composition of these genomes has also driven the conversion of many CAR codons to TAR once the original chain-terminating TARs had been fortuitously reduced in number to the point where release factors recognizing them could be lost. However, the genome of *Hexamita* appears to be GC rich: the overall and third position GC content of these genes from *Hexamita* 50330 are 53 and 63%, respectively, and from *H.inflata* are 52 and 64%, respectively. This argues that AT mutation pressure is not necessary to explain the appearance of TAA and TAG glutamine codons.

If directional mutation pressure is not a requirement (although it may contribute in other situations), then the answer might lie in the fact that canonical glutamine CAA and CAG codons and anticodons are, together with TTG tryptophan codons, the only sense triplets that can be converted to TAA or TAG by single transitions. Novel tRNA genes arising by chance duplication and base substitutions in the anticodon will not be maintained by selection until codons which require their services have also arisen by chance within coding regions. Both events will generally take place most frequently when they result from transitions rather than transversions (Kimura, 1980), so modifications to the specificity of TAR codons will tend to involve glutamine, regardless of directional mutation pressure.

Capture of TAA and TAG by glutamine could be further facilitated if G-U pairing would allow a single tRNA with anticodon UUA to recapture both UAA and UAG as Gln codons, and this has been suggested (Osawa *et al.*, 1992).

However, uridine residues in the first position of NNR decoding tRNAs are usually modified to one of several derivatives which bind strongly to A and weakly to G (Björk, 1995), necessitating a second tRNA to decode NNG. Even in *Tetrahymena*, where the first position U exhibits a rare modification which does allow both A and G to be recognized (Schull and Beier, 1994), there are still two tRNAs to decode the variant Gln codons UAA and UAG (Hanyu *et al.*, 1986). Based on the results described here, it appears that *Hexamita* 50330 and *H.inflata* use both isoacceptors to decode TAR. This may point to a deeper general reason why two tRNAs are always required to decode TAR, but it is simpler to suppose that species of *Hexamita* modify U34 in a more conventional fashion than *Tetrahymena*. This supposition is supported by the presence of $tRNA^{Gln}_{CUG}$ in *S.muris* which suggests that in other diplomonads both CUG and UUG isoacceptors are used to decode CAR.

Why TAA and TAG together? Without a tRNA which efficiently recognizes both codons, it is unlikely that the conversion of TAA and TAG to Gln codons took place simultaneously. A more plausible scenario is that two separate iterations of the codon capture process took place, each involving one unassigned codon. If this is correct, then the fact that in eukaryotes TAA and TAG are always reassigned together may mean that both have to be lost as functioning nonsense codons before either can be recaptured as sense.

If, for instance, some activity of the eukaryotic peptide release mechanism was common to termination exclusively by TAA and TAG, then neither of these codons could appear as sense within the coding regions of genes until that activity was rendered superfluous, and lost. This, in turn, could not take place until neither codon was absolutely essential for the termination of any gene. A possible role for release factors in this process is also suggested by the phylogenetic restriction of glutamine-specifying TAR codons to the eukaryotic nucleus. The eubacterial peptide termination system does not appear to be homologous to that of eukaryotes (Frolova *et al.*, 1994; Zhouravleva *et al.*, 1995) and uses two codon-specific factors, one recognizing UAA and UAG, and the other recognizing UAA and UGA (Caskey *et al.*, 1968; Scolnick *et al.*, 1968). This redundancy makes the loss of UAA (and because of wobble this extends to UAG) very unlikely in eubacteria. Indeed, only UGA has been lost in any eubacterial or organellar system, in contrast to the nucleus where changes involving UAA and UAG are the most common alteration to the nuclear genetic code.

Materials and methods

Culture conditions and nucleic acid extractions

ATCC 50380 was grown in Keister's Modified TYI-S-33 at 15°C in airtight 15 ml tubes. Cultures were maintained in dark, microaerophilic conditions for ~10 days before harvesting. Maximum cell density was very low, so a large number of cultures were combined before nucleic acids could be extracted. ATCC 50330 was grown under similar conditions by A.Roger, who provided frozen cells. Total DNA was isolated from *Hexamita* cells by harvesting up to 150 ml of culture by centrifugation, followed by repeated phenol and cetyltrimethylammonium bromide (CTAB) extractions.

Amplification of α-tubulin, β-tubulin and EF-1α from diplomonads

Primers designed for α-tubulin, β-tubulin and EF-1α were used in PCRs with genomic DNA templates from diplomonads. Products of the expected size were cloned using the TA cloning vector pCRII (Invitrogen) and sequenced using the T7 sequencing kit from Pharmacia. All PCR products were sequenced in duplicate on both strands, and three α-tubulin products from ATCC 50330 were sequenced from two independent amplifications to rule out error during amplification (no differences were observed). Primers used for the amplification of α-tubulin (TCCGAATTCARGTNGGAAYGCNTGYTGGGA and TCCAAGCTTCCATNCCYTCNCCNACRTACCA) and β-tubulin (TCCTCGAGTRAAYTCCATYTCRTCAT and TCCTGCAGGNCARTGYGGNAAYCA) were provided by A.Roger, and those for EF-1α (amplified in two overlapping pieces using CGAGGATCCGTTATTGGNCAYGTNGA and ACGTTGGATCCAACRTTRTCNCC for the 5′ end and GGTCGCGACAGTYTGNCTCATRTC with species-specific primers for the 3′ end) were provided by S.Baldauf.

Amplification of tRNA genes

Amplification reactions using primers based on known $tRNA^{Gln}$ sequences, GGTACCGGKYCYATGGYSTARTGGTA and GGTACCGGGCCYRSYSGGATTCGAAC, were used with total DNA from individual diplomonad species as template. All reactions were carried out separately on different occasions to avoid cross-contamination. In addition, for each species two reactions were performed on separate days, and clones from each sequenced and compared. Products were separated on polyacrylamide gels resulting in a band of the expected size (~84 bp) which was isolated according to the protocol of Maxam and Gilbert (1977), cloned and sequenced as described above.

Phylogenetic analysis

Conceptual translations of diplomonad EF-1α genes were aligned with homologues retrieved from existing databases and phylogenetic trees inferred. Parsimony trees were determined by conducting 50 random addition heuristic searches using PAUP (version 3.1.1; Swofford, 1993). Distance matrices based on the Dayhoff PAM 250 substitution matrix were calculated with PROTDIST and trees constructed by neighbour joining with NEIGHBOR programs in the PHYLIP package (version 3.5; Felsenstein, 1993). Statistical confidence was estimated for both methods by conducting 100 bootstrap replicates each. Maximum likelihood trees were determined using the PROTML program from the MOLPHY package (Adachi and Hasegawa, 1992). The computational complexity of maximum likelihood restricted the data set that could be examined, so the taxa were limited to three archaebacteria, all diplomonads and another deep-branching eukaryote, *Entamoeba histolytica*. The branching order of the archaebacteria was constrained to match the well-accepted division between crenarchaeota and euryarchaeota, and the branching order among diplomonads left completely unrestricted.

Acknowledgements

We thank H.van Kulen for DNA from *H.inflata* and *S.muris*, A.Roger for DNA from *G.lamblia*, A.Stoltzfus for discussion and S.Baldauf for help with MOLPHY. This work was supported by a grant (MT 4467) from the Medical Research Council of Canada. P.J.K. is a recipient of an MRC Studentship and W.F.D. is a Fellow of the Canadian Institute for Advanced Research.

References

Adachi,J. and Hasegawa,M. (1992) *Computer Science Monographs, No. 27. MOLPHY: Programs for Molecular Phylogenetics, I.—PROTML: Maximum Likelihood Inference of Protein Phylogeny.* Institute of Stastical Mathematics, Tokyo, Japan.

Björk,G.R. (1995) Biosynthesis and function of modified nucleosides. In Söll,D. and RajBhandary,U. (eds), *tRNA Structure, Biosynthesis and Function.* American Society for Microbiology, Washington, pp. 165–205.

Caskey,C.T., Tompkins,R., Scolnick,E., Caryk,T. and Nirenberg,M. (1968) Sequential translation of trinucleotide codons for the initiation and termination of protein synthesis. *Science*, **162**, 135–138.

Cavalier-Smith,T. (1993) The kingdom Protista and its 18 phyla. *Microbiol. Rev.*, **57**, 953–994.

Felsenstein,J. (1993) *PHYLIP (Phylogeny Inference Package).* University of Washington, Seattle, WA.

Frolova,L. *et al.* (1994) A highly conserved eukaryotic protein family possessing properties of polyeptide chain release factor. *Nature*, **372**, 701–703.

Hanyu,N., Kuchino,Y., Nishimura,S. and Beier,H. (1986) Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two Tetrahymena tRNAs Gln. *EMBO J.*, **5**, 1307–1311.

Hasegawa,M. and Fujiwara,M. (1993) Relative efficiencies of the maximum likelihood, maximum parsimony and neighbor joining methods for estimating protein phylogeny. *Mol. Phyl. Evol.*, **2**, 1–5.

Kimura,M. (1980) A simple method for estimating evolutionary rates of base substitution through comparitive studies of nucleotide sequences. *J. Mol. Evol.*, **17**, 110–113.

Leipe,D.D., Gunderson,J.H., Nerad,T.A. and Sogin,M.L. (1993) Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.*, **59**, 41–48.

Maxam,A.M. and Gilbert,W. (1977) A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA*, **74**, 560–564.

Osawa,S. and Jukes,T.H. (1989) Codon reassignment (codon capture) in evolution. *J. Mol. Evol.*, **28**, 271–278.

Osawa,S., Jukes,T.H., Watanabe,K. and Muto,A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.*, **56**, 229–264.

Prescott,D.M. (1994) The DNA of ciliated protozoa. *Microbiol. Rev.*, **58**, 233–267.

Rozario,C., Morin,L., Roger,A.J., Smith,M.W. and Müller,M. (1996) Primary structure and phylogenetic relationships of glyceraldehyde-3-phosphate dehydrogenase genes of free-living and parasitic diplomonad flagellates. *J. Eur. Microbiol.*, in press.

Schneider,S.U., Leible,M.B. and Yang,X.-P. (1989) Strong homology between the small subunit of ribose-1,5-bisphosphate carboxylase-oxygenase of two species of *Acetabularia* and the occurrence of unusual codon useage. *Mol. Gen. Genet.*, **218**, 445–452.

Schull,C. and Beier,H. (1994) Three Tetrahymena tRNA(Gln) isoacceptors as tools for studying unorthodox codon recognition and codon context effects during protein synthesis in vitro. *Nucleic Acids Res.*, **22**, 1278–1280.

Scolnick,E., Tompkins,R., Caskey,C.T. and Nirenberg,M. (1968) Release factors differing in specificity for terminator codons. *Proc. Natl Acad. Sci. USA*, **61**, 768–774.

Sprinzl,M., Hartmann,T., Weber,J., Blank,J. and Zeidler,R. (1989) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **17**(Suppl.), 1–173.

Swofford,D.L. (1993) *PAUP: Phylogenetic Analysis Using Parsimony.* Illinois Natural History Survey, Champaign, IL.

Tourancheau,A.B., Tsao,N., Klobutcher,L.A., Pearlman,R.E. and Adoutte,A. (1995) Genetic code deviations in the ciliates: evidence for multiple and independent events. *EMBO J.*, **14**, 3262–3267.

Zhouravleva,G., Frolova,L., Le Goff,X., Le Guellec,R., Inge-Vechtomov,S., Kisselev,L. and Philippe,M. (1995) Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J.*, **14**, 4065–4072.

Received on October 26, 1995; revised on January 10, 1996