doi:10.1016/S0022-2836(03)00057-3 *J. Mol. Biol.* (2003) **326**, 1337–1349

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Characterisation of a Non-canonical Genetic Code in the Oxymonad *Streblomastix strix*

Patrick J. Keeling* and Brian S. Leander

Department of Botany
Canadian Institute
for Advanced Research
University of British Columbia
3529-6270 University
Boulevard, Vancouver
BC, Canada V6T 1Z4

The genetic code is one of the most highly conserved characters in living organisms. Only a small number of genomes have evolved slight variations on the code, and these non-canonical codes are instrumental in understanding the selective pressures maintaining the code. Here, we describe a new case of a non-canonical genetic code from the oxymonad flagellate *Streblomastix strix*. We have sequenced four protein-coding genes from *S. strix* and found that the canonical stop codons TAA and TAG encode the amino acid glutamine. These codons are retained in *S. strix* mRNAs, and the legitimate termination codons of all genes examined were found to be TGA, supporting the prediction that this should be the only true stop codon in this genome. Only four other lineages of eukaryotes are known to have evolved non-canonical nuclear genetic codes, and our phylogenetic analyses of α-tubulin, β-tubulin, elongation factor-1 α (EF-1 alpha), heat-shock protein 90 (HSP90), and small subunit rRNA all confirm that the variant code in *S. strix* evolved independently of any other known variant. The independent origin of each of these codes is particularly interesting because the code found in *S. strix*, where TAA and TAG encode glutamine, has evolved in three of the four other nuclear lineages with variant codes, but this code has never evolved in a prokaryote or a prokaryote-derived organelle. The distribution of non-canonical codes is probably the result of a combination of differences in translation termination, tRNAs, and tRNA synthetases, such that the eukaryotic machinery preferentially allows changes involving TAA and TAG.

*Corresponding author

Keywords: genetic code; oxymonad; glutamine; translation; evolution

Introduction

The genetic code is the key to information flow in cellular life, and therefore one of the most highly conserved characteristics of all living systems. Virtually all genetic systems, including bacteria, archaebacteria, eukaryotic nuclei, organelles, and viruses use the same genetic code, which has been called the Universal genetic code. It is now widely accepted that this standard or canonical genetic code originated early in cellular evolution, prior to the common ancestor of all extant life. However, despite the widespread and near-absolute conservation of the genetic code, the code is not quite universal. A small number of genomes have been found to possess slightly different genetic codes, referred to here as non-canonical genetic codes. All non-canonical codes differ from the standard code only slightly, showing beyond any doubt that they have evolved by making small changes to the standard code. Various models have been suggested to explain how changes to the genetic code can be made, the most widely discussed being the codon-capture[1] and ambiguous intermediate[2] models. In codon-capture, shifts in the frequency of various codons, typically due to a strong bias in AT content of the genome, lead to the total loss of a particular codon. Once lost, the translational machinery responsible for recognising that codon can be lost or altered, leaving the codon "unassigned". Unassigned codons may then be reclaimed by the appearance of a mutant tRNA that can recognise them, which in turn allows the codon to reappear in the genome, potentially encoding a new amino acid.[1,3] The ambiguous

Abbreviations used: SSU, small subunit; EF, elongation factor; RT, reverse transcriptase; UTR, untranslated region; RACE, rapid amplification of cDNA ends; DIC, differential interference contrast.

E-mail address of the corresponding author: pkeeling@interchange.ubc.ca

intermediate model postulates that mutant tRNAs appear to claim active codons, for a time resulting in codons with two possible meanings. Eventually, the tRNA or termination factor originally recognising the codon is lost or altered so that it no longer recognises the codon, resulting in a new code.[2,4] While the order of events in these two models may be different, the basic mechanism is similar in some respects.

Although the differences between non-canonical genetic codes and the standard code are typically slight, these differences are important, since they reflect rare escapes from the strong selective pressures that constrain the code; namely, the activity of the protein translational machinery. To understand these pressures, it is necessary to have some appreciation for the variant codes that have evolved, and the frequency with which different non-canonical codes have appeared in different lineages. Mitochondrial genomes are a particularly fertile source of genetic code variations, likely because they are small genomes with unusual population dynamics.[5,6] In contrast, plastid genomes, which are similar in many respects to mitochondrial genomes, use the standard genetic code. Among eubacteria, a variant code has been documented in mycoplasmas,[7] but no variant code has been observed in any archaebacterium. In eukaryotic nuclear genomes, four lineages have been documented to alter their genetic code, and the distribution and nature of these changes is quite interesting. One unique variant is found in several species of the yeast Candida, where CUG encodes serine rather than leucine.[8] In ciliates, no less than three non-canonical genetic codes have evolved: in Euplotes, UGA encodes cysteine rather than stop,[9] while in Blepharisma and Colpoda, UGA encodes tryptophan,[10] and in most other ciliates TAA and TAG encode glutamine rather than stop.[11–13] This last variation is most interesting, since it appears to have evolved within ciliates more than once,[10,14] and has been documented in dasycladacean green algae[15] and in hexamitid diplomonads.[16,17] These non-canonical codes have yielded some intriguing insights into the stability of the genetic code in different lineages, but the number of variant codes characterised to date are still few. Therefore, the debates as to how these changes take place and how they may be a reflection of the translational apparatus have not abated.[18–20]

The oxymonads are a group of flagellates that rank among the most poorly studied eukaryotes known. They are anaerobes or microaerophiles that have no recognisable mitochondrion. The group is morphologically very diverse, but its members are not abundant in nature, being restricted to the guts of animals, predominantly termites.[21] Living in complex associations with other microorganisms and their animal hosts, oxymonads are not available in cultivation and accordingly it is only recently that molecular data have been characterised from any oxymonad.[22–24] The molecular data remain sparse and restricted to two closely related genera, so we sought to characterise several gene sequences from *Streblomastix strix*, an unusual oxymonad found in the hindgut of the damp-wood termite *Zootermopsis angusticolis*. Unexpectedly, *S. strix* was found to employ a non-canonical genetic code, where the amino acid glutamine is encoded by the canonical CAA and CAG codons, and by TAA and TAG codons.

This is only the fifth lineage of eukaryotes known to employ a non-canonical genetic code in the nuclear genome, but already the fourth lineage where this particular variant is found. The recurrent evolution of this variant code in nuclear genomes suggests that these codons are particularly free to evolve a new meaning in eukaryotes, and the different constraints on the genetic code in eukaryotes and prokaryotes likely reflect basic differences in translation.

Results and Discussion

Characterisation of molecular sequences from Streblomastix, and *in situ* hybridization

Fractionated *Z. angusticolis* hindgut material enriched with *S. strix* (see Figure 1(a) and (b) for morphology) was used to amplify several different gene sequences. Initially, the small subunit ribosomal RNA (SSU rRNA) gene was amplified to evaluate the proposed relationship between *S. strix* and oxymonads, as oxymonads are very poorly studied and morphologically diverse, and *S. strix* is a relatively unusual flagellate.[25–27] Amplifications of SSU rRNA genes from fractionated gut material yielded two products: one small (1574 bp), faint product corresponding to the size expected of parabasalia (which also are found in the *Z. angusticolis* hindgut), and a second, abundant and very large product (2471 bp). Both products were cloned and sequenced. The sequence of the smaller product corresponded to the recently reported sequence of *Trichomitopsis termopsidis*,[28] as expected, while the larger product corresponded to an SSU gene most similar to the large homologue (2012 bp) from the oxymonad Pyrsonympha. Preliminary phylogenetic analysis (not shown) confirmed this relationship. The putative *S. strix* SSU rRNA gene contains numerous large insertions, some at positions shared by insertions in the Pyrsonympha SSU, and others at unique positions.

To confirm the origin of the SSU rRNA gene, one unique insertion was used as a target for a *S. strix*-specific probe in fluorescent *in situ* hybridisation. When hybridised to whole gut contents, the probe specifically recognised *S. strix* (Figure 1(e) and (f)). Some wood-eating parabasalia were observed to fluoresce weakly at this wavelength, but these cells autofluoresced at the same intensity in controls that were not exposed to the

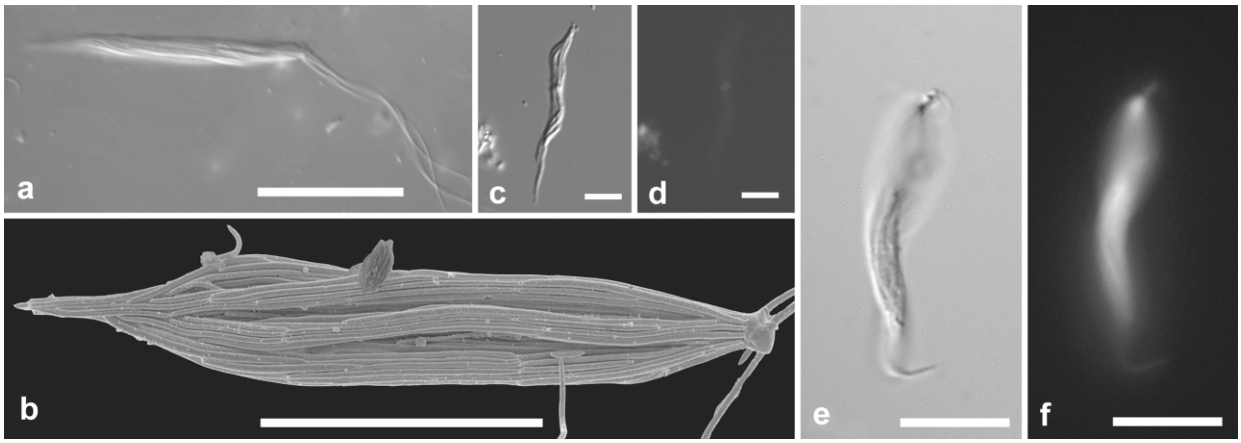

Figure 1. Representative light and scanning electron micrographs of *Streblomastix strix*; all scale bars represent 10 μm. (a) Light micrograph using differential interference contrast (DIC) showing the general appearance of a single cell. This appearance matches previous descriptions of *Streblomastix*.[26,27] (b) Scanning electron micrograph showing the rod-shaped epibiotic bacteria oriented longitudinally along the surface of a cell, a feature that is diagnostic of *S. strix*.[25] (c) DIC micrograph of a cell prepared for *in situ* hybridisation but without exposure to a rhodamine-labeled probe. (d) The same cell as in (c) observed under a 546 nm excitation wavelength; there was no evidence of autofluorescence. (e) DIC micrograph of a cell exposed to a rhodamine-labeled probe specific to *S. strix*. (f) Fluorescence emitted from the same cell as in (e) (under a 546 nm excitation wavelength) provides qualitative evidence for specific binding of the rhodamine-labeled SSU rDNA probe to *S. strix*.

rhodamine-labeled probe. In these controls, Streblomastix cells did not autofluoresce (Figure 1(c) and (d)). Moreover, in the preparations exposed to the probe, the degree of fluorescence in Streblomastix was noticeably higher than in parabasalids. As further confirmation, exact-match SSU rRNA primers were used to amplify the same SSU gene from manually isolated Streblomastix cells.

Fractionated gut material was used to amplify Streblomastix α-tubulin genes. A single product of the expected size was observed and three individual clones were sequenced. All three clones encoded distinct but similar α-tubulin genes that varied predominantly at synonymous positions. In preliminary phylogenetic analyses, these genes showed a very close relationship with the previously characterised α-tubulins from the oxymonads Pyrsonympha and Dinenympha, suggesting that the genes were indeed from Streblomastix. Moreover, the PCR-based approach using exact-match primers on 50 manually isolated Streblomastix cells produced a fragment of the same α-tubulin gene. Sequencing this product confirmed that these sequences are derived from Streblomastix.

The inferred translation of the PCR products revealed that all three copies of the α-tubulin that were sequenced encoded TAG codons at position 234, and two of the three copies encoded TAA codons at position 69. The third copy encoded a CAA glutamine codon at position 69 (Figure 2(a)). TAA and TAG are termination codons in the canonical genetic code, and both position 69 and 234 are highly conserved for glutamine in other eukaryotes, suggesting that TAA and TAG perhaps encode glutamine in the Streblomastix genome rather than acting as termination codons.

To determine whether this characteristic is common to other Streblomastix genes, genes encoding β-tubulin, translation elongation factor-1 α (EF-1 α), and heat-shock protein 90 (HSP90) were amplified, and several individual clones sequenced for each gene (five for β-tubulin and HSP90, and three for EF-1 α). As with α-tubulin, the individual clones for each of these genes encoded nearly identical proteins, but displayed a great deal of variation at synonymous positions (in HSP90, an extremely variable acidic domain was found to vary in sequence and length at the inferred amino acid level). Considering that the source of the material is a natural population and not a culture, this is perhaps not surprising: a natural population would be expected to have considerably more variability than normally seen in cultivated protozoa. As with α-tubulin, preliminary phylogenetic analysis confirmed that the Streblomastix EF-1 α branched with other known oxymonad genes (not shown). No other oxymonad β-tubulins or HSP90 genes are known, but oxymonads are known to be close relatives of the flagellate Trimastix.[22] Phylogenies of all four protein-coding genes including homologues from Trimastix (unpublished *Trimastix marina* sequences were made available by A. B. G. Simpson & A. J. Roger) showed Streblomastix and Trimastix to be closely related (not shown). To further confirm the origin of these sequences, Streblomastix was manually isolated once more, and fragments of β-tubulin and HSP90 were amplified from two pools of 50 isolated cells using exact-match primers. In each case, a product of the expected size was obtained and sequencing confirmed that the product matched the sequences of the Streblomastix genes.

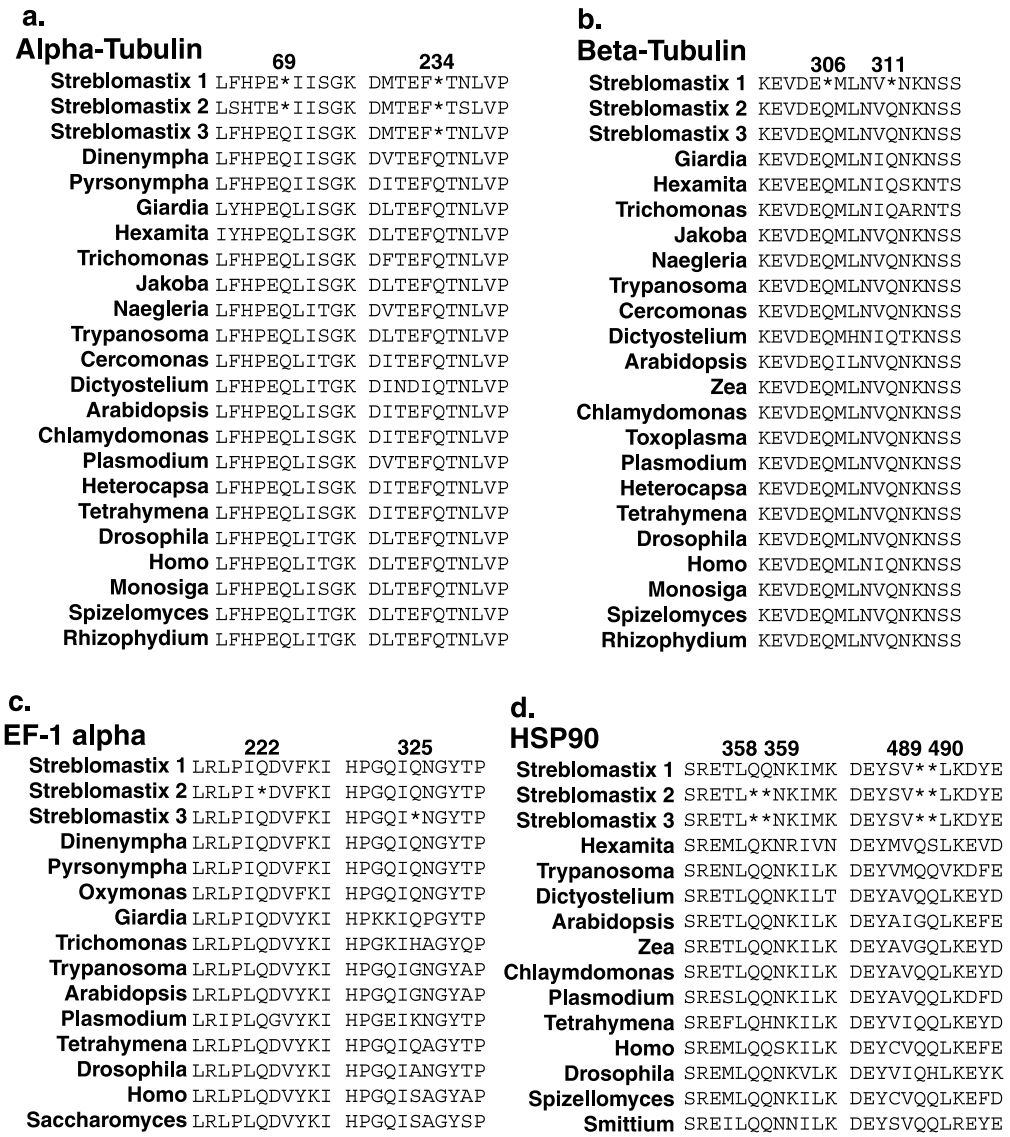

Figure 2. Examples of position and variability of TAA and TAG codons in *Streblomastix* genes. (a) α-Tubulin; (b) β-tubulin; (c) EF-1 α; (d) HSP90. Each section shows small blocks from a protein alignment with TAA or TAG codons in Streblomastix represented as an asterisk (*). Numbers above the sequence indicate the codon number in the Streblomastix sequence. In cases where different Streblomastix sequences encoded different codons at the position in question, all sequences are shown.

More importantly, every copy of each of these four genes encoded TAA and TAG codons at positions otherwise highly conserved for glutamine (Figure 2(b)–(d)). This confirms that the presence of TAA and TAG codons in α-tubulin is not gene-specific, and is a general feature of the genome. Indeed, a total of 116 TAA and TAG codons were observed at 32 positions throughout the various copies of the four genes, which is a substantial representation. Significantly, of the 32 positions where TAA or TAG codons were observed, CAA or CAG codons were used at the same position in one or more of the other copies of the gene in 26 cases (over 81% of the time), suggesting that these codons are interchangeable.

Characteristics of Streblomastix mRNA, termination codons, and 3′ UTRs

If Streblomastix uses a non-canonical genetic code where TAA and TAG encode glutamine, then all protein-coding genes in the genome are predicted to terminate with TGA codons. To test this, we performed 3′ rapid amplification of cDNA ends (RACE) on all four protein-coding genes. This method is extremely informative for two additional reasons. First, protein-coding gene sequences from oxymonads were, until now, restricted to α-tubulin and EF-1 α from two closely related genera, Pyrsonympha and Dinenympha, and an EF-1 α from an unidentified oxymonad. In total, these sequences include 24 positions

Table 1. Synonymous substitution patterns in Streblomastix

	α-Tubulin	β-Tubulin	EF-1 α	HSP90
Total TAR + CAR	12	19	19	18
Synonymous substitutions (observed:potential[a])	0.1845	0.3904	0.3657	0.2448
Non-synonymous substitutions (observed:potential[b])	0.0103	0.0066	0.0369	0.0205
YAA-YAG conversions (observed:potential)	0.2500	0.2632	0.3158	0.2222
TAR-CAR conversions (observed:potential)	0.1667	0.3158	0.5789	0.3889

[a] The frequency of an amino acid multiplied by the number of positions within its codons that can change resulting in a synonymous substitution (see Materials and Methods).
[b] The frequency of an amino acid multiplied by the number of positions within its codons that can change resulting in a non-synonymous substitution (see Materials and Methods).

encoding glutamine, and in none of these has a TAA or TAG codon been observed, suggesting that these oxymonads use the canonical genetic code. However, every one of these genes was acquired by reverse transcriptase (RT)-PCR from mRNA, leaving open the formal, albeit unlikely, possibility that oxymonads use a T-to-C RNA editing mechanism, and that we are observing pre-edited genes in Streblomastix. Second, general characteristics of protein carboxy termini and mRNA 3′ untranslated regions (UTRs) hold information about potentially important characteristics of the genome, and whether different transcripts are from different alleles or loci.

For each of the four genes, 3′ RACE products of the expected size were amplified, cloned, and five individual clones sequenced. As with the genomic clones, a substantial degree of synonymous variation was observed between different RACE clones, and between the RACE clones and the genomic PCR products, such that only three redundant sequences were found. In β-tubulin and HSP90, all 3′ UTRs were clearly distinct but related, while the 3′ UTRs of α-tubulin and EF-1 α fell into two distinct classes that shared no detectable sequence similarity. This suggests that the variation observed in coding regions might reflect different alleles (those with similar UTRs) as well as different loci (those with different UTRs). In either case, each distinct RACE product is clearly derived from a unique coding sequence, and the termination codon of each product is accordingly evolving independently. It is significant, therefore, that every one of the 16 unique RACE products terminated at the expected position with a TGA codon (many genes terminate with two TGA codons spaced 3 bp apart). In addition, all RACE products contained TAA and TAG codons at positions otherwise conserved for glutamine, and the proportion of CAR to TAR matches closely the pattern observed in genomic sequences (2:1). Altogether, there is no indication of RNA editing. This suggests that this non-canonical code is either restricted to Streblomastix, or that Pyrsonympha and Dinenympha use the non-canonical codons at a much lower frequency. To determine the distribution of this code in oxymonads, it will be necessary to characterise the 3′ ends of protein-coding genes from Pyrsonympha and Dinenympha, and to characterise more protein-coding genes from diverse oxymonads. Given the poor sampling of molecular data from oxymonads, it is reasonable to suspect that this character may be widespread among species not yet examined at the molecular level.

Codon usage in Streblomastix

Comparing the sequences of both genomic DNAs and mRNAs from all copies of all four protein-coding genes reveals that virtually every amplification product characterised is unique (nearly all of the variation being synonymous). The 3′ UTR sequences suggest that most of this variation can be attributed to variation between alleles, since in most cases the 3′ UTR sequences for a given gene are clearly homologous. In EF-1 α and α-tubulin mRNAs, however, two entirely non-homologous classes of 3′ UTR were observed. This suggests that, in at least these two instances, two distinct loci were characterised.

If TAA and TAG encode glutamine, then TAR/CAR variation should closely resemble other synonymous variation (this would not be observed if there was a very strong bias toward one or the other, but this is not the case). Taking a representative amino acid sequence for each gene, the potential number of synonymous and non-synonymous substitutions can be calculated (Table 1), and compared with the actual number observed between those copies of the gene that have been sequenced. For each gene, these ratios are similar (some difference being attributable to the number of copies of the gene that have been characterised: see Materials and Methods), and the difference between the two is about an order of magnitude (Table 1, rows 2 and 3). If substitutions at both the first and third positions of glutamine codons are examined, the ratios are found to be in line with ratios calculated for synonymous positions for that gene (Table 1, rows 4 and 5), suggesting that C-T substitutions at the first position are synonymous.

If the variation in codon use between different alleles and loci are considered to be at least partly independent, a table of codon frequencies for 30,804 bp of coding sequence can be compiled from all unique sequences. While this does not

represent the codon biases throughout the Streblomastix genome, because this is a skewed set of genes (three being constitutively highly expressed), it does reveal some trends. First, there is an obvious bias in favour of the canonical CAR glutamine codons over the non-canonical TAR codons, as they appear at a ratio of approximately 2:1. There is no consistent AT or GC bias (for instance, asparagine codons are biased to C at the third position, aspartate codons are biased to T). Similarly, there is no consistent purine or pyrimidine bias (for instance, there is an A bias in serine, proline, threonine, and alanine and a T bias in valine). The most conspicuous bias is towards A residues at the third position of codon quartets. While it is true that existing sequences cannot reveal all past mutation biases in a genome, the existing data from Streblomastix reveal no strong bias that could be interpreted as a driving force to change the genetic code.

Phylogeny and the distribution of non-canonical genetic codes in eukaryotes

Traditionally,[29,30] and more recently,[31] oxymonads have most often been considered to be related to retortamonads and diplomonads. Recent molecular data have suggested that this is not the case[22–24] but, nevertheless, the occurrence of the same non-canonical genetic code in an oxymonad and the hexamitid diplomonads[16,17] does require some careful attention to determine whether the two variant codes really evolved independently. To this end, phylogenetic trees were inferred using the five genes characterised to determine the distribution of this code among eukaryotes. Trees based on the four protein-coding genes are shown in Figure 3, and a phylogeny of SSU rRNA genes is shown in Figure 4, where all lineages with non-canonical nuclear genetic codes are indicated. In all cases where other oxymonad sequences are known (SSU rRNA, α-tubulin, and EF-1 α), Streblomastix branches with the other oxymonads, as expected. In no case does Streblomastix or the oxymonads in general show a specific relationship to any other lineage that uses this or any other non-canonical code. In particular, no phylogeny shows a close relationship between Streblomastix and the hexamitid diplomonads, although the β-tubulin phylogeny does show Streblomastix as a close relative of the parabasalia and diplomonads as a whole. Even in this case, the two lineages sharing the same divergent code are separated by the parabasalia and the diplomonad *Giardia intestinalis*, both of which use the standard code.[16,17] Altogether, the phylogenies do not show strong support for any particular position for Streblomastix in the eukaryotic tree, other than being related to other oxymonads and Trimastix. Regardless of the precise evolutionary origin of Streblomastix and the oxymonads, the phylogenies demonstrate unambiguously that the non-canonical code in Streblomastix evolved independently of other non-canonical codes, including that found in diplomonads.

The evolution of the genetic code

Figure 4 shows an SSU rRNA phylogeny including all eukaryotic lineages known to use a non-canonical genetic code, and indicates how each differs from the standard code. It is difficult to determine exactly how many times the code has changed in eukaryotes because the phylogeny of ciliates is uncertain and the code has changed many times in ciliates (likely a result of ciliates adopting their unique system of differentiated germ-line and somatic nuclei[32]), but it is clear that the code has changed in five distinct lineages. It is interesting that in four of these lineages, the same non-canonical code has appeared, and there are strong reasons to believe it has appeared more than once independently in ciliates,[10,14] while the other three codes known in nuclear genomes are each found only once (although a good case can be made for two origins of UGA encoding tryptophan in Blepharisma and Colpoda[10]).

Code changes involving stop codons are detected far more easily than changes where codons shift from encoding one amino acid to another, so it is possible that these dominate our understanding of code changes simply because additional changes have gone undetected in poorly studied genomes. However, there are reasons to believe that stop codons could be reassigned more easily than other codons. Most importantly, stop codons are rare in any genome (each represented a maximum of once per gene, and more likely much less), and the fewer the number of codons in question the easier it is to reassign them, since it demands fewer events and fewer potentially deleterious intermediates. Nevertheless, it would appear that TAA and TAG are especially prone to reassignments in eukaryotic genomes, and when they are reassigned, they always encode glutamine. Why such a bias would exist, and why it would be phylogenetically restricted are interesting questions, and the answer is probably a complex mix of the various factors that constrain the genetic code. We will consider three such factors: the translation termination apparatus, tRNAs and tRNA synthetases, and mutation frequencies.

First, the translation termination systems of eubacteria, archaebacteria, and eukaryotes have now been characterised partially, and there are substantial differences between eubacterial and nuclear systems. In particular, eubacteria use two termination factors: RF1 recognises UAA and UAG, and RF2 recognises UAA and UGA.[33–35] In contrast, eukaryotes and archaebacteria possess a single protein factor, eRF1, which recognises all three stop codons.[36–38] It has been argued that changes to the codon recognition site of eukaryotic eRF1 can lead to the loss of UAA and UAG recognition.[10,39] Conversely, it is probably very difficult to reassign UAA in eubacteria, since there

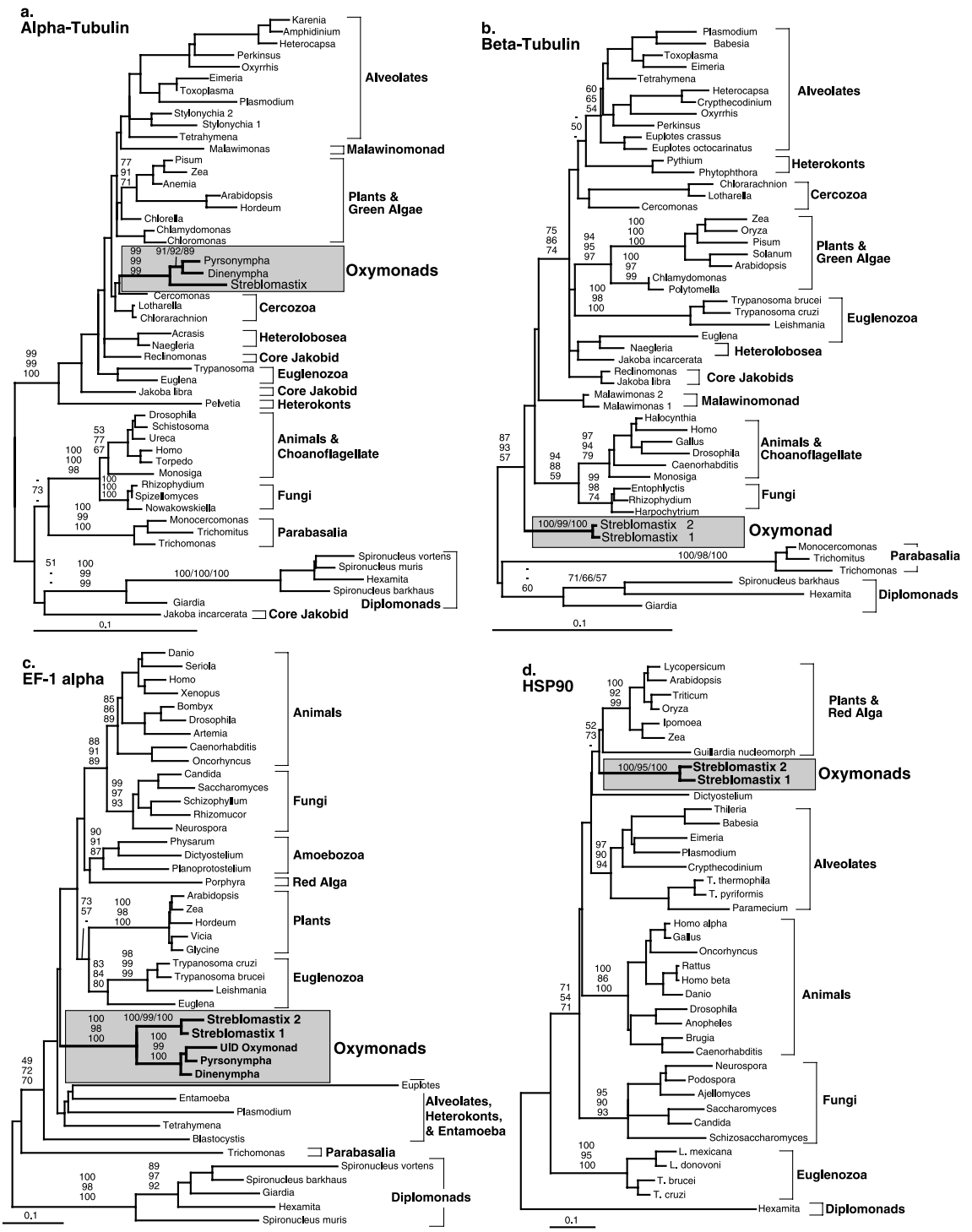

Figure 3. Phylogenetic trees of protein-coding genes. (a) α-Tubulin; (b) β-tubulin; (c) EF-1 α; (d) HSP90. All trees are unrooted γ-corrected, weighted neighbor-joining trees (diplomonads are shown arbitrarily as the outgroup for consistency, and so the relative position of diplomonads and oxymonads can be seen easily). Numbers at nodes correspond to bootstrap support from weighted neighbor-joining (top), Fitch-Margoliash (centre), and protein maximum likelihood (bottom). Major eukaryotic groups are bracketed and named to the right.

are two independent factors that recognise this codon. This could, partly, explain why UAA and UAG have never been reassigned together in eubacteria or eubacterial organelles (although UAG has been reassigned alone as either alanine or leucine in the mitochondria of various green algae[40]).

Second, a critical part of ensuring the fidelity of the genetic code is in the specificity of tRNAs and the enzymes that charge them with the appropriate amino acid, the tRNA synthetases. In virtually every system that has been examined, CAR glutamine codons are decoded using two tRNAs, one with anticodon CUG and one with UmUG (where

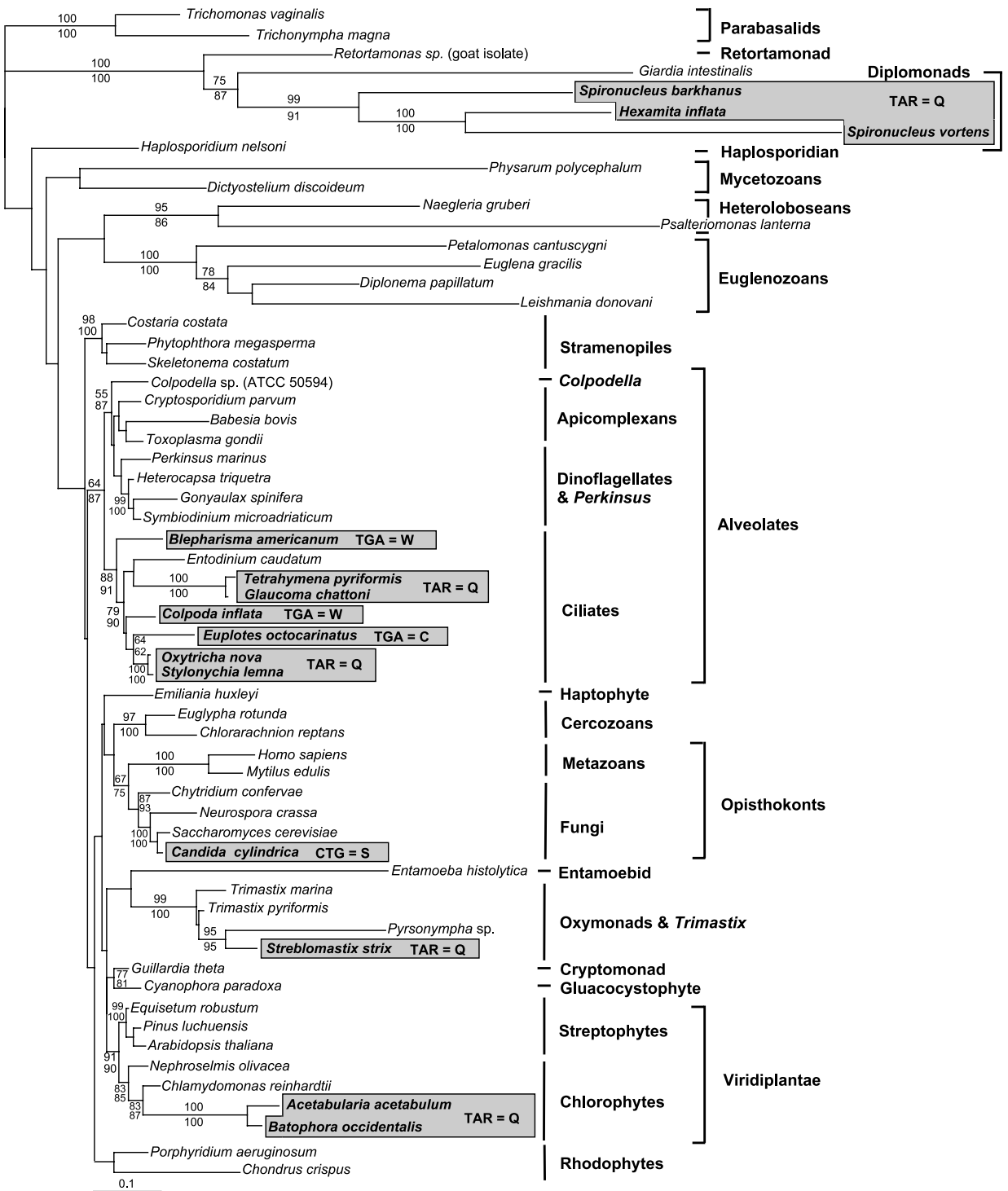

Figure 4. The distribution of non-canonical genetic codes in eukaryotes. γ-Corrected ML tree ($-\ln L = 25353.473$) inferred from an alignment containing 60 SSU rDNA sequences and 1140 sites showing the position of Streblomastix among the oxymonads and other eukaryotic lineages with non-canonical genetic codes (all boxed). Numbers at relevant nodes correspond to 100 bootstrap replicates using γ-corrected weighted neighbor-joining (above) and Fitch-Margoliash (below). Major eukaryotic lineages are labeled to the right. The transition/transversion ratio was 1.69; the scale bar represents 0.1 substitution (corrected) per site.

Um is a modified uridine base). Interestingly, both isoacceptors have been shown to act as natural supressors of UAG and UAA in various eukaryotes,[41,42] providing a natural bias for UAR codons being reassigned to glutamine.

While the fidelity of the code is generally thought of at the level of tRNA binding to mRNA, it is equally important that the tRNA be charged with the correct amino acid. The recognition of the appropriate tRNA by its cognate aminoacyl-tRNA

synthetase takes place through the so-called identity elements of the tRNA.[43] In most cases (including glutamine), the anticodon makes up an important part of the identity set, so one necessary step in the origin of a non-canonical genetic code is the relaxation of the tRNA-specificity of the tRNA synthetase. In most cases, there is a single tRNA synthetase specific for each amino acid, but there are some exceptions, and glutamine is one of note. Glutaminyl-tRNAs are charged by a glutaminyl-tRNA synthetase in all known nucleo-cytoplasmic systems of eukaryotes, but are otherwise found only in certain eubacteria, where they are considered to have originated by lateral gene transfer.[44] All other eubacteria, eubacterial-derived organelles, and archaebacteria lack a specific glutaminyl-tRNA synthetase, and instead charge glutaminyl-tRNAs with glutamic acid using the glutamyl-tRNA synthetase.[43–45] The glutamate residue is subsequently transamidated by a specific Glu-tRNA^{Gln} amidotransferase,[46] yielding glutaminyl-tRNA. It is conceivable that the specific recognition of tRNA^{Gln} by not one, but two enzymes in these organisms (the tRNA synthetase and the amidotransferase) could limit the probability of charging mutant tRNAs in these eubacteria, organelles, and archaebacteria. Taking together the translation termination apparatus, the presence of natural suppressor tRNAs and a discrete glutaminyl-tRNA synthetase, the eukaryotic nucleo-cytosolic environment could be a unique intersection of several factors that favour this particular non-canonical genetic code.

Lastly, however, the natural variation in the frequencies of different kinds of mutations should be considered, as these too could play an important role in the evolution of non-canonical genetic codes. A single transition is by far the most common mutation, and six different codons can be converted into TAA or TAG by single transitions. A transition mutation at the first position (C-to-T) converts glutamine codons CAA and CAG to TAA and TAG. At the second position, a single A-to-G transition converts the lone tryptophan codon (TGG) to TAG, while the TGA-to-TAA mutation is a synonymous opal-to-ochre change. At the third position, TAA and TAG mutate to one another, and are therefore also synonymous. Accordingly, if TAG and TAA are to be involved in a codon reassignment, it is most likely to involve either glutamine or tryptophan. A change to tryptophan is considerably less likely than to glutamine, since tryptophan is the rarest amino acid: for a codon reassignment to be fixed in the genome, it is necessary for canonical codons to mutate to the new codon, and there are relatively few tryptophan codons in a genome available to do this. Overall, the impact of mutation frequencies might be slight, but the potentially frequent mutation between TAR and CAR, together with the presence of an additional synonymous codon (ochreTGA) makes the TAR pair especially flexible, and perhaps prone to reassignment when other factors are also favourable.

The exact course of events that leads to the evolution of a non-canonical genetic code has been modeled in several different ways, and many attempts have been made to take into account the nature of the genomes involved, the tRNAs that result, and the effects on translation termination factors to come up with a unified explanation for how the genetic code evolves.[3–6,10,18–20,39,47,48] However, the evolution of the genetic code might better be regarded as a balance between many factors: in this case these may include the nature of the translation termination equipment, characteristics of tRNAs, the system used to charge glutaminyl-tRNAs, and natural biases in mutation frequencies. Indeed, it might be more informative to treat each novel genetic code individually, since each new code has evolved under a unique set of circumstances, potentially for very different reasons and by very different means. There is no reason to believe that a single explanation or a single model for the evolution of the genetic code exists, but rather a set of outcomes whose frequencies reveal important aspects of the nature of the translational machinery, mutation, and the flexibility of the code.

Materials and Methods

Collection, identification, and isolation of *Streblomastix strix*

The termite *Z. angusticolis* was collected from damp and rotting wood at Jerico Beach, Vancouver, BC, Canada. Termites were maintained in the laboratory in plastic containers with damp wood from their original collection site. Termite hindgut contents were diluted in Trager's medium U[49] and inspected by differential interference contrast (DIC) light microscopy for the presence of Streblomastix, which was identified in all termites examined. As the hindgut of *Z. angusticolis* contains several other flagellates (but only one oxymonad), the material was enriched with Streblomastix by a simple density fractionation. Hindgut contents from ten termites were drawn into a 10 cm glass Pasteur pipette and allowed to settle for five minutes. The contents of the pipette were expelled slowly into three equal fractions, and the process repeated three times on the last fraction from each round. Microscopic examination revealed that this material was predominantly Streblomastix, while the remainder was largely the parabasalian flagellate Trichomitopsis, and a very small number of large trichonymphid parabasalia. Fractionated hindgut material was harvested by centrifugation for the extraction of DNA and RNA. DNA was extracted by resuspending pelleted material in 100 μl of CTAB extraction buffer (1.12 g of Tris, 8.18 g of NaCl, 0.74 g of EDTA, 2 g of cetyltrimethylammonium bromide (CTAB), 2 g of polyvinylpyrolidone, 0.2 ml of 2-mercaptoethanol in 100 ml of water) in a Knotes Duall 20 tissue-homogeniser and incubating at 65 °C with periodic grinding. The lysate was extracted twice with chloroform/isoamyl alcohol (24:1, v/v), and the aqueous phase precipitated in 95% (v/v) ethanol.

Individual cells of Streblomastix were also manually isolated from gut contents diluted in Trager's medium, using a Pasteur pipette stretched to a diameter of approximately 75 μm. Isolated cells were re-diluted in clean medium and re-isolated a total of three times until no contaminating eukaryotes could be observed. DNA was prepared from isolated cells by extracting homogenised cells with chloroform/isoamyl alcohol (24:1, v/v) and precipitation in ethanol. PCR-based approaches using exact-match primers (described later) were used to confirm that gene sequences were derived from the genome of Streblomastix.

Amplification and sequencing

The SSU rRNA genes from Streblomastix were amplified from DNA extracted from the fractionated hindgut material using the primers TGATCCTTCTGCAGGT TCACCTAC and CTGGTTGATCCTGCCAGT. Protein-coding genes from Streblomastix could not be amplified using a simple protocol, and were amplified using a two-step nested PCR procedure. In each case a reaction was carried out using a pair of outer primers on the DNA extracted from the fractionated hindgut material, and the product of that reaction used as the template for a secondary reaction using a pair of inner primers. Primers used were: for α-tubulin GGGCCCCAGGTCGG CAAYGCNTGYTGG and GGGCCCCGAGAACTCSC CYTCYTCCAT (outer primers) and CGCGGCCTCARG TNGGNAAYGCNTGYTGGGA and CGCGCCATNCCY TCNCCNACRTACCA (inner primers); for β-tubulin GCCTGCAGGNCARTGYGGNAAYCA and TCCTCGA GTRAAYTCCATYTCRTCCAT (outer primers) and CAG ATCGGCGCGAARTTYTGGGARAT and CTCGTCCAT GCCYTCNCCNGTRTACCA (inner primers); for EF-1 α AACATCGTCGTGATHGGNCAYGTNGA and CTTGAT CACNCCNACNGCNACNGT (outer primers) and CA ACATCGTCGTCATCGGNCAYGTNGA and GCCGCG CACGTTGAANCCNACRTTRTC (inner primers); and for HSP90 GGAGCCTGATHATHAAYACNTTYTA and CGCCTTCATDATNCKYTCCATRTTNGC (outer primers) and ACGTTYTAYWSNAAYAARGARAT and CGCCTT CATDATNCKYTCCATRTTNGC (inner primers). HSP90 sequences were also obtained using the inner primers ACGTTYTAYWSNAAYAARGARAT and GATGACYTT NARDATYTTRTTYTGYTG. The HSP90 gene from *Hexamita inflata* was also amplified using GGAGCCTGATH ATHAAYACNTTYTA and CGCCTTCATDATNCKYTC CATRTTNGC so that a representative diplomonad could be included in the phylogeny of HSP90. For each gene, several independent clones were sequenced completely.

PCR-based experiments using exact-match primers on manually isolated cells of Streblomastix were used to help confirm that the clones were not derived from a different eukaryotic genome. Exact-match primers (available on request) were designed to amplify small fragments (400–500 bases) of the SSU rRNA, α-tubulin, β-tubulin and HSP90 genes. Amplified products were cloned, sequenced in one direction, and compared to the original sequences.

Microscopy and *in situ* hybridisation

Cells observed with DIC light microscopy were suspended in Trager's medium, fixed in 1% (w/v) glutaraldehyde, and secured under a cover-slip with VALAP (vaseline/lanolin/paraffin; 1:1: by wt[50]). Images were produced with a Zeiss Axioplan 2 Imaging microscope connected to a Q-Imaging, Microimager II, black and white digital camera.

Cells for scanning electron microscopy were prepared with an osmium vapor fixation protocol.[51] A small volume (10 ml) of cells suspended in Trager's medium was transferred into a Petri dish that contained a piece of filter-paper mounted on the inner surface of the lid. The filter-paper was saturated with 4% (w/v) OsO_4 and the lid placed over the dish exposing the cells to OsO_4 vapors. After 30 minutes of exposure, the cells were fixed for an additional 30 minutes with six drops of 4% OsO_4 added directly to the medium. Cells were transferred onto a 3 μm polycarbonate membrane filter (Corning Separations Div., Acton, MA), dehydrated with a graded series of ethyl alcohol, and critical point dried with CO_2. Filters were mounted on stubs, sputter-coated with gold, and viewed under a Hitachi S4700 scanning electron microscope (SEM). The SEM image was presented on a black background using Adobe Photoshop 6.0 (Adobe Systems, San Jose, CA).

In situ hybridisation using a rhodamine-labeled oligonucleotide probe was conducted to help demonstrate that the PCR-amplified SSU rDNA sequence was derived from the genome of Streblomastix. The probe consisted of 24 bases taken from an insertion unique to the putative Streblomastix rRNA gene: Rhodamine-CTATTGGT CATCAGCTGCAGTGCG (t_m in 1 M Na^+ = 76 °C). Gut contents of *Z. angusticolis* were suspended in Trager's medium, pelleted with gentle microcentrifugation (8 rpm), and pre-fixed in 4% paraformaldehyde in 5 × Tris buffer (2.5 M NaCl, 0.1 M Tris, pH 7.5) for 30 minutes. Following dehydration in a graded series of ethyl alcohol, cells were gradually rehydrated using 3:1, 1:1, 1:3, and 0:1 ratios of ethyl alcohol and TBSt buffer (Tris buffer, 0.1% (v/v) Tween 20). Cells were partially digested with proteinase K for 15 minutes and washed three times with TBSt (ten minutes each). Cells were post-fixed with 4% paraformaldehyde in TBSt buffer.

A pre-hybridisation step involved washing the cells three times (ten minutes each) in a hybridisation mix (5 × SSC, 0.2% Tween 20, 5 mM EDTA). Cells were heated in the hybridisation mix at 70 °C for 45 seconds before the fluorescent-labeled probe was added. Once the probe was added (3 μl of probe to cells suspended in 120 μl of the hybridisation mix), cells were immediately exposed to a denaturing step for five minutes at 95 °C followed by an annealing step for five minutes at 60 °C. Unbound probe was washed from the cells five times (ten minutes each) with the hybridisation mix. Rhodamine-labeled cells were viewed under a Zeiss Axioplan 2 Imaging microscope using a 546 nm excitation wavelength (emission wavelength 590 nm). Cells were also prepared as described, but without exposure to the rhodamine-labeled probe, to examine autofluorescence (i.e. false positives). Negative controls for non-specific binding were available by examining the degree of fluorescence in parabasalids such as Trichomitopsis and Trichonympha in the same preparations.

Characterisation of mRNAs by 3′ RACE

RNA was isolated from fractionated Streblomastix (see above) by resuspending harvested material in 1 ml of Trizol (Gibco-BRL) and transferring it to a Knotes Duall 20 tissue homogeniser. Material was ground for five minutes and incubated for five minutes at room temperature without grinding. Lysate was extracted with 200 μl of chloroform/isoamyl alcohol (24:1, v/v),

and the aqueous phase precipitated with 500 μl of isopropanol. RT-PCR was carried out using 600 ng of total RNA as a template with the M-MLV reverse transcriptase and poly(T) adapter primer from the RLM-RACE kit (Ambion). 3′ RACE was performed on the first-strand DNA using the common anchor primer GCGAGCACA GAATTAATACGACT together with gene-specific primers: CTTTGTTAGAGCACACTGATGTTGC for α-tubulin, GCTTCAAGACACTCAAGCTGTCGAACC for β-tubulin, ATGGATATACACCAGTGCTTGATTG or TGTGGAGATTCTAAATAAGATCCAC for EF-1 α, and CAGTTCAAACATCTCCATTCTTAGA or TCGTGTTCT TATTATGGAGAATTGCG for HSP90. Reactions were carried out according to the RLM-RACE protocol using AmpliTaq gold polymerase (Applied Biosystems). In all cases, a product was recovered after the first round of amplification, cloned, and sequenced as above.

Phylogenetic and codon analysis

The synonymous nature of first-position mutations at glutamine codons was examined by two simple ratios. For each gene, a representative (consensus) amino acid sequence was taken, and the number of potential synonymous and non-synonymous substitutions was calculated according to the following formulae. Potential non-synonymous substitutions are the frequency of each amino acid multiplied by the number of positions within its codons that yield a non-synonymous change when mutated (these factors are: A,G,L,P,R,T,V = 2; C,D,E,F,H, I,K,M,N,S,W,Y = 3). Potential synonymous substitutions are the frequency of each amino acid multiplied by the number of positions within its codons that yield a synonymous change when mutated (these factors are: M,W = 0; A,C,D,E,F,G,H,I,K,N,P,T,V,Y = 1; L,R = 2; S = 3). The number of observed synonymous and non-synonymous substitutions are counted using the variation observed between different copies of each gene sequenced from DNA or cDNA. This is an oversimplification, since it does not take into account multiple substitutions, and the number is an arbitrary result of the number of alleles and loci characterised (e.g. if more copies of α-tubulin were characterised, the number of potential substitutions would remain the same, but the number observed would have to increase). This has little effect on the conclusion, however, since the actual value of the ratio is not important, but rather the relationship of the value for glutamine codons compared with that of all other codons.

To infer the phylogenetic position of Streblomastix genes, representative sequences were aligned with homologues from public databases and phylogenetic trees inferred using distance and maximum likelihood methods. For protein data, distances were inferred using TREE-PUZZLE 5.0[52] with the WAG substitution matrix, and site-to-site rate variation modeled on a γ-distribution with eight rate categories, and the α parameter and fraction of invariable sites estimated from the data. The estimated shape parameter (α) and proportion of invariable sites (*i*) were: for α-tubulin, $\alpha = 0.40$ and $i = 0$; for β-tubulin, $\alpha = 0.43$ and $i = 0$; for EF-1 α, $\alpha = 0.53$ and $i = 0$; and for HSP90, $\alpha = 0.70$ and $i = 0$. Trees were constructed using weighted neighbor-joining with WEIGHBOR 1.0.1a,[53] and Fitch-Margoliash using FITCH 3.6a.† Bootstraps were carried out by the same methods using PUZZLEBOOT (shell script available from www.tree-puzzle.de). Protein maximum likelihood trees were inferred using ProML 3.6a with global rearrangements and ten random addition replicates. Site-to-site rate variation was modeled using the -R option, with invariable sites and six categories of variable sites estimated by TREE-PUZZLE. Protein ML bootstrapping was carried out under the same parameters except that four categories of variable sites were considered. SSU rRNA phylogenies were inferred by maximum likelihood using PAUP 4.0b10‡ with the HKY substitution frequency matrix using a heuristic search starting with a neighbor-joining tree and site-to-site rate variation modeled as above. The α-parameter and proportion of invariable sites were estimated by TREE-PUZZLE ($\alpha = 0.43$ and $i = 0$). Bootstraps were calculated in the same way. Distance trees and distance bootstraps were inferred using the same parameters.

Data Bank accession numbers

All new sequences have been deposited in GenBank under accession numbers AY138769-AY138805.

Acknowledgements

This work was supported by a grant (227301-00) from the Natural Sciences and Engineering Research Council of Canada (NSERC). P.J.K. is a scholar of the Canadian Institute for Advanced Research, the Michael Smith Foundation for Health Research, and the Canadian Institutes for Health Research. B.S.L. is supported by a postdoctoral fellowship from the National Science Foundation (USA). We thank Alastair Simpson, Jeff Silberman, and Andrew Roger for access to unpublished sequences from Trimastix, and J. M. Archibald, J. T. Haper, and an anonymous reviewer for critical reading of the manuscript.

References

1. Osawa, S. & Jukes, T. H. (1989). Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **28**, 271–278.
2. Schultz, D. W., Yarus, M. & Transfer, R. N. A. (1994). mutation and the malleability of the genetic code. *J. Mol. Biol.* **235**, 1377–1380.
3. Osawa, S., Jukes, T. H., Watanabe, K. & Muto, A. (1992). Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**, 229–264.
4. Schultz, D. W. & Yarus, M. (1996). On malleability in the genetic code. *J. Mol. Evol.* **42**, 597–601.
5. Knight, R. D., Landweber, L. F. & Yarus, M. (2001). How mitochondria redefine the code. *J. Mol. Evol.* **53**, 299–313.
6. Kurland, C. G. (1992). Evolution of mitochondrial genomes and the genetic code. *BioEssays*, **14**, 709–714.
7. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. & Osawa, S. (1985). UGA is

† http://evolution.genetics.washington.edu/phylip/doc/fitch.html

‡ http://paup.csit.fsu.edu/problems.html

read as tryptophan in *Mycoplasma capricolum*. *Proc. Natl Acad. Sci. USA*, **82**, 2306–2309.
8. Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. & Iwasaki, S. (1989). The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. *Nature*, **341**, 164–166.
9. Meyer, F., Schmidt, H. J., Plumper, E., Hasilik, A., Mersmann, G., Meyer, H. E. *et al.* (1991). UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus*. *Proc. Natl Acad. Sci. USA*, **88**, 3758–3761.
10. Lozupone, C. A., Knight, R. D. & Landweber, L. F. (2001). The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.* **11**, 65–74.
11. Caron, F. & Meyer, E. (1985). Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature*, **314**, 185–188.
12. Helftenbein, E. (1985). Nucleotide sequence of a macronuclear DNA molecule coding for alpha-tubulin from the ciliate *Stylonychia lemnae*. Special codon usage: TAA is not a translation termination codon. *Nucl. Acids Res.* **13**, 415–433.
13. Horowitz, S. & Gorovsky, M. A. (1985). An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc. Natl Acad. Sci. USA*, **82**, 2452–2455.
14. Baroin-Tourancheau, A. B., Tsao, N., Klobutcher, L. A., Pearlman, R. E. & Adoutte, A. (1995). Genetic code deviations in the ciliates: evidence for multiple and independent events. *EMBO J.* **14**, 3262–3267.
15. Schneider, S. U., Leible, M. B. & 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.
16. Keeling, P. J. & Doolittle, W. F. (1996). A non-canonical genetic code in an early diverging eukaryotic lineage. *EMBO J.* **15**, 2285–2290.
17. Keeling, P. J. & Doolittle, W. F. (1997). Widespread and ancient distribution of a non-canonical genetic code in diplomonads. *Mol. Biol. Evol.* **14**, 895–901.
18. O'Sullivan, J. M., Davenport, J. B. & Tuite, M. F. (2001). Codon reassignment and the evolving genetic code: problems and pitfalls in post-genome analysis. *Trends Genet.* **17**, 20–22.
19. Ribas de Pouplana, L. & Schimmel, P. (2001). Aminoacyl-tRNA synthetases: potential markers of genetic code development. *Trends Biochem. Sci.* **26**, 591–596.
20. Syvanen, M. (2002). Recent emergence of the modern genetic code: a proposal. *Trends Genet.* **18**, 245–248.
21. Brugerolle, G. & Lee, J. J. (2000). Order Oxymonadida. In *The Illustrated Guide to the Protozoa* (Lee, J. J., Leedale, G. F. & Bradbury, P., eds), 2nd edit., pp. 1186–1195, Society of Protozoologists, Lawrence, KA.
22. Dacks, J. B., Silberman, J. D., Simpson, A. G., Moriya, S., Kudo, T., Ohkuma, M. & Redfield, R. J. (2001). Oxymonads are closely related to the excavate taxon *Trimastix*. *Mol. Biol. Evol.* **18**, 1034–1044.
23. Moriya, S., Tanaka, K., Ohkuma, M., Sugano, S. & Kudo, T. (2001). Diversification of the microtubule system in the early stage of eukaryote evolution: elongation factor 1 alpha and alpha-tubulin protein phylogeny of termite symbiotic oxymonad and hypermastigote protists. *J. Mol. Evol.* **52**, 6–16.
24. Moriya, S., Ohkuma, M. & Kudo, T. (1998). Phylogenetic position of symbiotic protist *Dinenympha* [correction of *Dinemympha*] *exilis* in the hindgut of the termite *Reticulitermes speratus* inferred from the protein phylogeny of elongation factor 1 alpha. *Gene*, **210**, 221–227.
25. Hollande, A. & Carruette-Valentin, J. (1970). La lignee des pyrsonymphines et les caracteres infrastructuraux communs aux genres *Opisthomitus*, *Oxymonas*, *Saccinobaculus*, *Pyrsonympha*, et *Streblomastix*. *Compt. Rend. Acad. Sci. ser. D*, **270**, 1587–1590.
26. Kidder, G. W. (1929). *Streblomastix strix*, morphology and mitosis. *Univ. Calif. Publ. Zool.* **33**, 109–124.
27. Kofoid, C. A. & Swezy, O. (1919). Studies on the parasites of the termites. I. On *Streblomastix strix*, a polymastigote flagellate with a linear plasmodial phase. *Univ. Calif. Publ. Zool.* **20**, 21–40.
28. Keeling, P. J. (2002). Molecular phylogenetic position of *Trichomitopsis termopsidis* (Parabasalia) and evidence for the Trichomitopsiinae. *Eur. J. Protistol.* **38**, 279–286.
29. Brugerolle, G. (1991). Flagellar and cytoskeletal systems in amitochondrial flagellates Archamoebae, Metamonada and Parabasalia. *Protoplasma*, **164**, 70–90.
30. Brugerolle, G. & Taylor, F. J. R. (1977). Taxonomy, cytology and evolution of the Mastigophora. In *Proceedings of the Fifth International Congress of Protozoology* (Hutner, S. H., ed.), pp. 14–28, Pace University, New York.
31. Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol. Rev.* **73**, 203–266.
32. Cohen, J. & Adoutte, A. (1995). Why does the genetic code deviate so easily in ciliates? *Biol. Cell.* **85**, 105–108.
33. Caskey, C. T., Tompkins, R., Scolnick, E., Caryk, T. & Nirenberg, M. (1968). Sequential translation of trinucleotide codons for the initiation and termination of protein synthesis. *Science*, **162**, 135–138.
34. Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. (1968). Release factors differing in specificity for terminator codons. *Proc. Natl Acad. Sci. USA*, **61**, 768–774.
35. Craigen, W. J., Lee, C. C. & Caskey, C. T. (1990). Recent advances in peptide chain termination. *Mol. Microbiol.* **4**, 861–865.
36. Dontsova, M., Frolova, L., Vassilieva, J., Piendl, W., Kisselev, L. & Garber, M. (2000). Translation termination factor aRF1 from the archaeon *Methanococcus jannaschii* is active with eukaryotic ribosomes. *FEBS Letters*, **472**, 213–216.
37. Frolova, L., Goff, X. L., Rasmussen, H. H., Cheperegin, S., Drugeon, G., Kress, M. *et al.* (1994). A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature*, **372**, 701–703.
38. Inagaki, Y. & Doolittle, W. F. (2000). Evolution of the eukaryotic translation termination system: origins of release factors. *Mol. Biol. Evol.* **17**, 882–889.
39. Inagaki, Y. & Doolittle, W. F. (2001). Class I release factors in ciliates with variant genetic codes. *Nucl. Acids Res.* **29**, 921–927.
40. Hayashi-Ishimaru, Y., Ohama, T., Kawatsu, Y., Nakamura, K. & Osawa, S. (1996). UAG is a sense codon in several chlorophycean mitochondria. *Curr. Genet.* **30**, 29–33.
41. Beier, H. & Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucl. Acids Res.* **29**, 4767–4782.
42. Grimm, M., Nass, A., Schull, C. & Beier, H. (1998). Nucleotide sequences and functional characterization of two tobacco UAG suppressor tRNA(Gln)

isoacceptors and their genes. *Plant Mol. Biol.* **38**, 689–697.
43. Ibba, M. & Soll, D. (2000). Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**, 617–650.
44. Brown, J. R. & Doolittle, W. F. (1999). Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *J. Mol. Evol.* **49**, 485–495.
45. Freist, W., Gauss, D. H., Ibba, M. & Soll, D. (1997). Glutaminyl-tRNA synthetase. *Biol. Chem.* **378**, 1103–1117.
46. Curnow, A. W., Hong, K., Yuan, R., Kim, S., Martins, O., Winkler, W. *et al.* (1997). Glu-tRNAGln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA*, **94**, 11819–11826.
47. Lehman, N. (2001). Molecular evolution: please release me, genetic code. *Curr. Biol.* **11**, R63–R66.
48. Moreira, D., Kervestin, S., Jean-Jean, O. & Philippe, H. (2002). Evolution of eukaryotic translation elongation and termination factors: variations of evolutionary rate and genetic code deviations. *Mol. Biol. Evol.* **19**, 189–200.
49. Trager, W. (1934). The cultivation of a cellulose-digesting flagellate. *Trichomonas termopsidis*, and of certain other termite protozoa. *Biol. Bull.* **66**, 182–190.
50. Kuznetsov, S. A., Langford, G. M. & Weiss, D. G. (1992). Actin-dependent organelle movement in squid axoplasm. *Nature*, **356**, 722–725.
51. Leander, B. S., Witek, R. P. & Farmer, M. A. (2001). Trends in the evolution of the euglenid pellicle. *Evolution*, **55**, 2215–2235.
52. Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969.
53. Bruno, W. J., Socci, N. D. & Halpern, A. L. (2000). Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.* **17**, 189–197.

Edited by J. Karn

(Received 24 October 2002; received in revised form 2 January 2003; accepted 3 January 2003)