How did UGA codon translation as tryptophan evolve in certain ciliates? A critique of Kachale et al. 2023 *Nature*

Estienne Carl Swart[1,*], Christiane Emmerich[1], Kwee Boon Brandon Seah[1], Minakshi Singh[1], Yekaterina Shulgina[2], Aditi Singh[1]

1. Max Planck Institute for Biology, Tuebingen, Germany
2. Department of Molecular and Cellular Biology, Harvard University, Cambridge, United States

* Corresponding author: estienne.swart@tuebingen.mpg.de

Abstract

Ciliates are a widespread clade of microbial eukaryotes with the greatest diversity of nuclear genetic codes (at least eight) following a recent addition[1]. All non-standard ciliate genetic codes involve stop codon reassignments[1,2,3]. Two of these codes are ambiguous[1–3], with “stop” codons either translated or terminating translation depending on their context[2,3]. Ambiguous genetic codes have arisen not only in ciliates, but also independently in trypanosomatids from the genus *Blastocrithidia*[4] and an alveolate species from the genus *Amoebophrya*[5]. Two ambiguous genetic codes in ciliates share translation of UGA “stop” codons as tryptophan with *Blastocrithidia* and the *Amoebophrya* species. tRNA genes with complementary anticodons to reassigned UAA and UAG stop codons have invariably been found in ciliate species that translate these codons[1,2]. Furthermore, though a UGA-cognate $tRNA^{Cys}_{UCA}$ was reported in *Euplotes*[6], a ciliate genus that translates UGA as cysteine, vexingly, no nuclear genome-encoded $tRNA^{Trp}_{UCA}$ has been found in ciliate species with UGA tryptophan codons. Recently, Kachale et al. provided evidence for UGA translation as tryptophan in *Blastocrithidia nonstop* and the ciliate *Condylostoma magnum* using 4 base pair anticodon stem (AS) near-cognate tryptophan $tRNA^{Trp}_{CCA}$’s, rather than the typical 5 base pair stem tRNAs[7]. New tRNA data we report from additional ciliates bolsters this hypothesis. Kachale et al. also hypothesised that a particular amino acid substitution in the key stop codon recognition protein, eRF1 (eukaryotic Release Factor 1), favours translation of UGA as tryptophan instead of termination[7]. Contrary to Kachale et al, we propose such substitutions favouring reduced eRF1 competition enhancing “stop” codon translation do not need to occur concomitantly with tRNA alterations or acquisitions to evolve new genetic codes via stop codon reassignment. We report multiple instances of the substitution investigated in Kachale et al. 2023 that have not led to UGA translation, and multiple ciliate species with UGA tryptophan translation but without the substitution, indicating it is not necessary. Consistent with the ambiguous intermediate hypothesis for genetic code evolution, experimental evidence and our observations suggest continued potential ciliate eRF1-tRNA competition.

Critique

There are two important issues in Kachale et al. 2023 to resolve from the outset, both clarified in Supporting Information. One: a reported 5 bp AS tryptophan tRNA (in their Fig. 3a) which did not lead to efficient stop codon readthrough[7], likely originates from a bacterium present in the ciliate culture, not from *Condylostoma magnum*. Thus, *Condylostoma* probably only has nuclear genome-encoded 4 bp AS $tRNA^{Trp}_{CCA}$'s, further supporting the proposal that they are necessary for efficient UGA translation as tryptophan. Two: the authors express uncertainty about the genetic codes used by several ciliate species, particularly relating to UGA codons and tRNAs with potential complementary anticodons; this can partly be traced to incorrect annotations in reference databases or earlier publications, that have since been superseded. This is pertinent to the interpretation of eRF1 substitutions and their role in UGA translation.

With their diverse genetic codes, ciliates are ideal for exploring hypotheses about these codes. Consistent with Kachale et al's 4 bp AS tryptophan tRNA proposal for translation of UGA as tryptophan, we find such tRNAs occur not only in *Condylostoma magnum* (class Heterotrichea; UAR=Q/*, UGA=W/*)[2,3] but also in another ciliate species with an ambiguous genetic code, *Loxodes magnus* (class Karyorelictea; UAG=Q, UAA=Q/*? [UAA may also be a stop], UGA=W/*)[8], and in the ciliate genus *Blepharisma* (class Heterotrichea; Fig. 1), which appears to translate UGA unambiguously as tryptophan (UAR=*, UGA=W)[8]. A draft *Loxodes magnus* somatic genome assembly has over one hundred 4 bp AS $tRNA^{Trp}$ genes but just six 5 bp AS $tRNA^{Trp}$ genes (Fig. 1). The 5 bp AS tRNAs of *L. magnus* shown in Fig 1. have one or two unpaired T-stem bases and are co-located on a contig with similar sequences and secondary structures, particularly the T-stem, but different anticodons (CUU, CUA, CCU) corresponding to non-tryptophan codons. This suggests relaxation of selection, and that these are likely pseudogenes. In contrast, two other heterotrich ciliates with standard genetic codes[9,10], *Stentor coeruleus* and *Fabrea salina*, only have 5 bp AS $tRNA^{Trp}$ genes (Fig. 1). Among members of the *Paramecium aurelia* complex, which are not known to translate UGA as tryptophan (code: UAR=Q, UGA=*)[2], we found a ciliate species, *Paramecium biaurelia*, with at least one 4 bp AS tRNA (Extended Data Fig. 1a) among its eight $tRNA^{Trp}$ genes. For such tRNAs and additional ones in other organisms that are not evident pseudogenes, the ability to translate UGA as tryptophan should be carefully experimentally investigated in future.

In a human pathogenic trypanosomatid *Leishmania* species, tRNA C-to-U editing of the wobble (5') anticodon base of nuclear genome-encoded mitochondrial $tRNA^{Trp}_{CCA}$'s generates UCA anticodons that enable mitochondrial UGA codon translation[11]. Kachale et al. reported no editing of cytosolic *Blastocrithidia* $tRNA^{Trp}_{CCA}$'s[7] that could permit translation of cytosolic UGAs. Previously, tRNA sequencing did not reveal C-to-U editing of the *Condylostoma magnum* $tRNA^{Trp}_{CCA}$ anticodon wobble base that would generate an anticodon complementary to UGA codons[2]. We have also observed no appreciable C-to-U editing (> 0.1% of tryptophan tRNAs) in *Blepharisma* and *Loxodes* 4 bp AS $tRNA^{Trp}$ sequences (in 172,929 and 9,721 unique reads, respectively; Supplementary Table 1, Source Data Fig. 1).

Anticipating the discovery of natural 4 bp AS $tRNA^{Trp}_{CCA}$'s in *Blastocrithidia* and the ciliates *Condylostoma*, *Blepharisma*, and *Loxodes*, the idea that tRNA AS mutations can enhance decoding of near-cognate codons was previously explored in back-to-back papers by Schultz and Yarus[12,13]. In the first paper, by extensive mutational screening of *Escherichia coli* tRNA su7 G36, a derivative of a tryptophan suppressor tRNA with a CUG anticodon, they found that mutations that disrupt the top AS stem base pair — creating a 4 bp AS stem — led to the most translation of UAG[12], which involves G-U wobble pairing at the 1st codon position that is normally disallowed. In the second paper, using su7 $tRNA_{CUA}$ they showed that mutations the disrupt the top AS stem base pair promote UAA translation[13], which involves mismatched 3rd position A-C like that required for UGA translation by 4bp AS $tRNA_{CCA}$'s[7].

Particular amino acid substitutions in homologs of the protein that recognizes stop codons, eRF1, were formerly thought to be associated with loss of stop codon recognition necessary for the evolution of particular genetic codes in ciliates[14,15]. With the benefit of additional eRF1 sequences and ciliate genetic codes, we previously reported multiple counterexamples to such associations[2]. Kachale et al. proposed that a single amino acid substitution in eRF1, from Ser67 to Ala/Gly67 (numbered with respect to yeast eRF1), may be needed for loss of UGA termination in conjunction with a shorter $tRNA^{Trp}$ anticodon stem for efficient UGA translation as tryptophan[7]. However, this substitution is present in eRF1's of multiple ciliates that use the standard genetic code: *Stentor coeruleus*, *Fabrea salina* and *Climacostomum virens* (all members of class Heterotrichea; like *Condylostoma* and *Blepharisma*; Fig. 2a). eRF1 of the ciliate *Pseudocohnilembus persalinus* (class Oligohymenophorea), which has the genetic code UAR=Q and UGA=*, also has Ala67, and so too does eRF1 of the diplomonad flagellate *Giardia intestinalis* (standard genetic code).

Furthermore, all karyorelict ciliates translate UGA as tryptophan within mRNA coding sequences (and use it as a stop at the ends of coding sequences)[8]. One species, *Loxodes magnus*, has eRF1's with Ala67, but other species' eRF1's have either Ser67 or Cys67 (Fig. 2a). Recently the ciliate species *Plagiopyla frontata* (class Plagiopylea) was reported to have ambiguous UGA codons that are translated as tryptophan in coding sequences (genetic code UAR=Q, UGA=W/*)[1]. eRF1 from this distantly related ciliate has Ser67. So too does eRF1 from the alveolate *Amoebophrya* sp. ex *Karlodinium venificum*. Thus, the Ala/Gly67 substitution is present in ciliate species without UGA translation and is not necessary in multiple ciliate species which translate UGA as tryptophan. Ala/Gly67 also appears unnecessary in the *Amoebophrya* species.

Unlike bacteria which have two proteins that recognize two stop codons each, RF1 and RF2, standard genetic code model eukaryotes, like yeast, typically have a single "omnipotent" protein, eRF1, that recognizes all three stop codons. eRF1 paralogs were previously noted to have arisen independently in certain ciliate genera, including *Euplotes* and *Tetrahymena*[16,17,18]. The frequent occurrence of such paralogs in ciliates raises the possibility some may have subfunctionalized, like RF1 and RF2, with different stop codon recognition capabilities, but this needs experimental determination. *Blepharisma stoltei* has three divergent eRF1 paralogs (60-72% amino acid identity for the three pairwise comparisons), of which the most highly

transcribed one (BSTOLATCC_MAC3627; mean 540 RPKM, standard deviation 50 RPKM for a developmental time series[19]) has Ala67, but there is also an eRF1 paralog with low transcription (mean 9.9 RPKM, standard deviation 6.2 RPKM) and Ser67 (Fig. 2a). Gene expression of the more divergent *Blepharisma* eRF1 paralogs is comparable to that of the ancient eRF1 paralog Dom34/Pelota (BSTOLATCC_MAC12938; mean 6.8 RPKM; standard deviation 2.3 RPKM), a protein responsible for the translation-associated process "No-Go decay"[16]. It is conceivable that these paralogs have functionally diverged, now serving an alternative role like Dom34/Pelota.

Though we have only observed UGA codons translated as tryptophan in *Blepharisma stoltei*, *in vitro* translation experiments suggest *Blepharisma japonicum*'s ortholog of the highly transcribed *B. stoltei* eRF1, also with Ala67, can recognize all three stop codons, but UGA the most weakly[20]. Correspondingly, with some capacity of *Blepharisma* eRF1 to recognize UGA as stop codons, a signal of potential competition between *B. stoltei* eRF1 and $tRNA^{Trp}$ can be observed in the form of UGA codon depletion in a region beginning 25-30 codons upstream of UAR stop codons (Fig. 2b). A similar depletion was observed in the karyorelict ciliates and the heterotrich ciliate *Condylostoma* which do use UGA as a stop close to transcript ends[2,8]. Interestingly, depletion of UAA and UGA codons occurs in a similar region before stops in *Blastocrithidia nonstop*, contrasting with the constancy of reassigned UAG codons[7]. This suggests eRF1-tRNA competition not only for UAA but also for UGA in this species.

While the type of amino acid substitution proposed by Kachale et al. may certainly substantially enhance translation, it should be noted that such substitutions are not a prerequisite for the acquisition of a new genetic code under the hypothesis which best fits the evolution of the ambiguous stop/sense genetic codes, the "ambiguous intermediate hypothesis"[21]. Instead, in a transitional evolutionary phase, codons may be interpreted in two ways, with potential eRF1-tRNA competition. With time, beneficial mutations or modifications in either the tRNA or eRF1 (or other components of translation) that reduce competition may be selected.

Instead of focusing on individual eRF1 substitutions, we propose future investigations should more generally explore the structure of non-standard genetic code eRF1's captured in translation termination in the context of their own ribosomes. New genetic codes involving stop codon reassignments have had ample opportunity to evolve in eukaryotes through a combination of tRNA and eRF1 mutations, but are limited to just a few clades, most notably having radiated in ciliates. We thus infer that an additional aspect has enabled genetic code evolution in these prolific microbes, and continue to suggest that this may be their ability to either tolerate or resolve genetic code ambiguity.

References

1. McGowan, J. *et al.* Identification of a Non-Canonical Ciliate Nuclear Genetic Code Where UAA and UAG Code for Different Amino Acids. *BioRxiv* (2022) doi:10.1101/2022.12.16.520718.
2. Swart, E. C., Serra, V., Petroni, G. & Nowacki, M. Genetic codes with no dedicated stop codon: Context-dependent translation termination. *Cell* **166**, 691–702 (2016).
3. Heaphy, S. M., Mariotti, M., Gladyshev, V. N., Atkins, J. F. & Baranov, P. V. Novel Ciliate Genetic Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons in *Condylostoma magnum*. *Mol. Biol. Evol.* **33**, 2885–2889 (2016).
4. Záhonová, K., Kostygov, A. Y., Ševčíková, T., Yurchenko, V. & Eliáš, M. An Unprecedented Non-canonical Nuclear Genetic Code with All Three Termination Codons Reassigned as Sense Codons. *Curr. Biol.* **26**, 2364–2369 (2016).
5. Bachvaroff, T. R. A precedented nuclear genetic code with all three termination codons reassigned as sense codons in the syndinean *Amoebophrya* sp. ex *Karlodinium veneficum*. *PLoS ONE* **14**, e0212912 (2019).
6. Turanov, A. A. *et al.* Genetic code supports targeted insertion of two amino acids by one codon. *Science* **323**, 259–261 (2009).
7. Kachale, A. *et al.* Short tRNA anticodon stem and mutant eRF1 allow stop codon reassignment. *Nature* **613**, 751–758 (2023).
8. Seah, B. K. B., Singh, A. & Swart, E. C. Karyorelict ciliates use an ambiguous genetic code with context-dependent stop/sense codons. *Peer Community Journal* **2**, (2022).
9. Slabodnick, M. M. *et al.* The macronuclear genome of *Stentor coeruleus* reveals tiny introns in a giant cell. *Curr. Biol.* **27**, 569–575 (2017).
10. Zhang, B. *et al.* An Extremely Streamlined Macronuclear Genome in the Free-Living Protozoan Fabrea salina. *Mol. Biol. Evol.* **39**, (2022).
11. Alfonzo, J. D., Blanc, V., Estévez, A. M., Rubio, M. A. & Simpson, L. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in Leishmania tarentolae. *EMBO J.* **18**, 7056–7062 (1999).
12. Schultz, D. W. & Yarus, M. tRNA structure and ribosomal function. I. tRNA nucleotide 27-43 mutations enhance first position wobble. *J. Mol. Biol.* **235**, 1381–1394 (1994).
13. Schultz, D. W. & Yarus, M. tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. *J. Mol. Biol.* **235**, 1395–1405 (1994).
14. Lehman, N. Molecular evolution: Please release me, genetic code. *Curr. Biol.* **11**, R63-6 (2001).
15. Lozupone, C. A., Knight, R. D. & Landweber, L. F. The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.* **11**, 65–74 (2001).
16. Atkinson, G. C., Baldauf, S. L. & Hauryliuk, V. Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol. Biol.* **8**, 290 (2008).
17. Inagaki, Y. & Doolittle, W. F. Class I release factors in ciliates with variant genetic codes. *Nucleic Acids Res.* **29**, 921–927 (2001).
18. Liang, H., Wong, J. Y., Bao, Q., Cavalcanti, A. R. O. & Landweber, L. F. Decoding the decoding region: analysis of eukaryotic release factor (eRF1) stop codon-binding residues. *J. Mol. Evol.* **60**, 337–344 (2005).
19. Swart, E. Blepharisma stoltei ATCC 30299 Macronuclear genome. *Edmond* (2022) doi:10.17617/3.8c.
20. Eliseev, B. D. *et al.* Translation termination factor eRF1 of the ciliate *Blepharisma japonicum* recognizes all three stop codons. *Mol Biol (NY)* **45**, 614–618 (2011).
21. Schultz, D. W. & Yarus, M. On malleability in the genetic code. *J. Mol. Evol.* **42**, 597–601 (1996).

Contributions

Conceptualization: E.C.S. Investigation: E.C.S., B.K.B.S, M.S. Methodology: E.C.S, C.E., B.K.B.S, A.S. Resources: all authors. Writing: all authors. Supervision: E.C.S.

Competing interests

The authors declare no competing interests.

Figure captions

Fig. 1. Predicted tryptophan tRNAs encoded in the macronuclear genomes of heterotrich and karyorelict ciliates. The two ciliate classes are indicated by different background colours: Heterotrichea - green; Karyorelictea - cyan. Nucleotide substitutions that differ between tRNA genes are indicated by double arrows. Stop codon reassignments are given under the species names. See Extended Data Fig. 1 for *Blepharisma*'s $tRNA_{UCA}$ predictions (selenocysteine and mitochondrial). *Blepharisma japonicum* and *Blepharisma undulans* $tRNA^{Trp}_{CCA}$ also have 4 bp anticodon stems (Source Data Fig. 1 and Extended Data Fig. 1).

Fig. 2. eRF1 substitutions and potential signals of eRF1-tRNA competition in *Blepharisma*. (a) eRF1 coordinates are given according to that of *Saccharomyces cerevisiae*, and the alignment window is the same as that in Kachale et al. 2023. The complete eRF1 alignment along with the sources of the sequences is provided in Source Data Fig. 2. For the genetic codes of each species, stars indicate stop codons, and question marks indicate possible stop codons. Check marks and crosses respectively indicate agreements and disagreements with respect to the proposed UGA assignment/eRF1 substitution at position 67. The eRF1 phylogeny to the left was generated by RAxML. The second *Geleia acuta* eRF1 paralog is encoded by an incomplete transcript. (b) Codon frequency upstream of predicted *B. stoltei* stop codons for the permutations of "T", "G" and "A" bases. For the complete codon frequency matrix, see Extended Data Fig. 2a.

Extended Data Fig. 1. Secondary structures of *Paramecium biaurelia* 4 bp AS $tRNA^{Trp}_{CCA}$ and *Blepharisma* $tRNA_{UCA}$'s and a tRNA-like molecule with a possible UCA anticodon. (a) *Paramecium biaurelia* $tRNA^{Trp}_{CCA}$; terminal nucleotides predicted by tRNAscan-SE 2.0. (b) *B. stoltei* $tRNA^{Sec}_{UCA}$. (c) *B. stoltei* mitochondrial $tRNA^{Trp}_{UCA}$. (d) tRNA-like molecule with potential UCA anticodon in *B. stoltei*. (e) Multiple sequence alignment of tRNA-like sequences from *Blepharisma* spp. (f) Multiple sequence alignment of paralogs of tRNA-like sequence paralogs in *B. stoltei* ATCC30299. (g) YAMAT-seq mapping to the *B. stoltei* ATCC30299 mitochondrial genome.

Extended Data Fig. 2. Codon usage before stops and base frequencies around stops. (a) Codon frequency upstream of predicted *B. stoltei* stop codons. (b) Base frequencies flanking *B. stoltei* TGA codons (n=44087). (c) Adenosine frequency of +4 at base immediately downstream of translated *B. stoltei* codons. (d) Base frequencies around TAA (n=21002) stop codons. (e) Base frequencies around TAG (n=4707) stop codons. (f) Base frequencies around TGG tryptophan codons (n=78535).

Fig 1

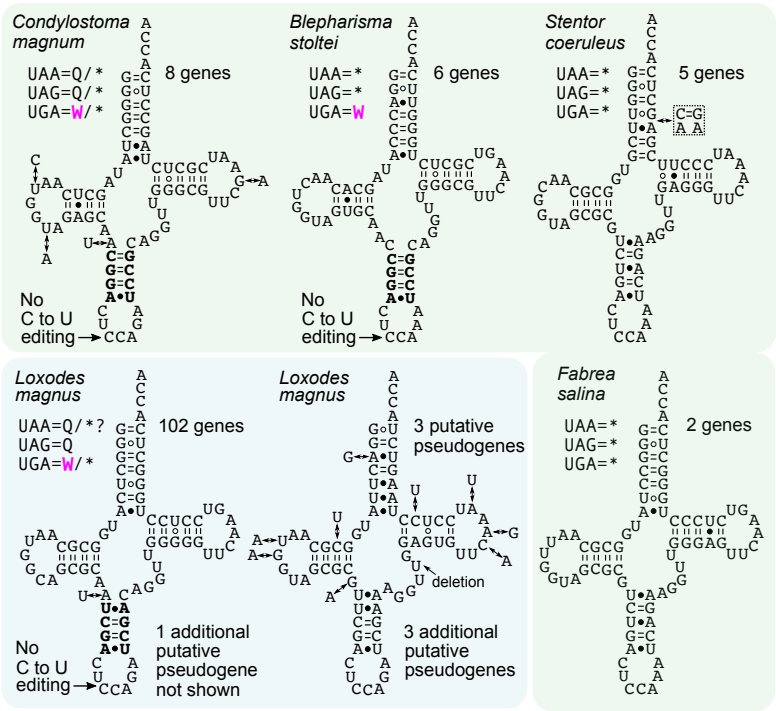
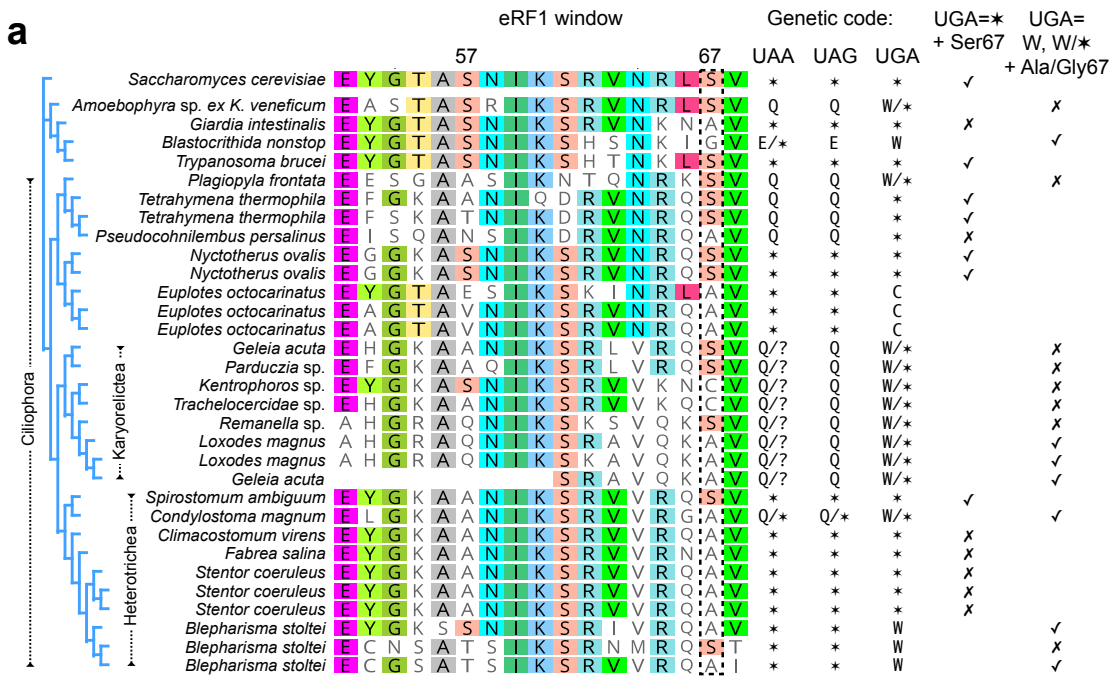
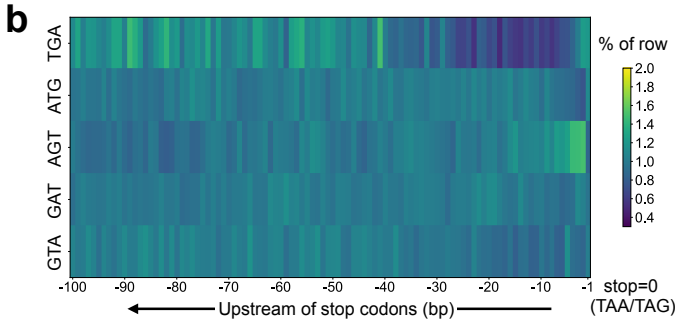
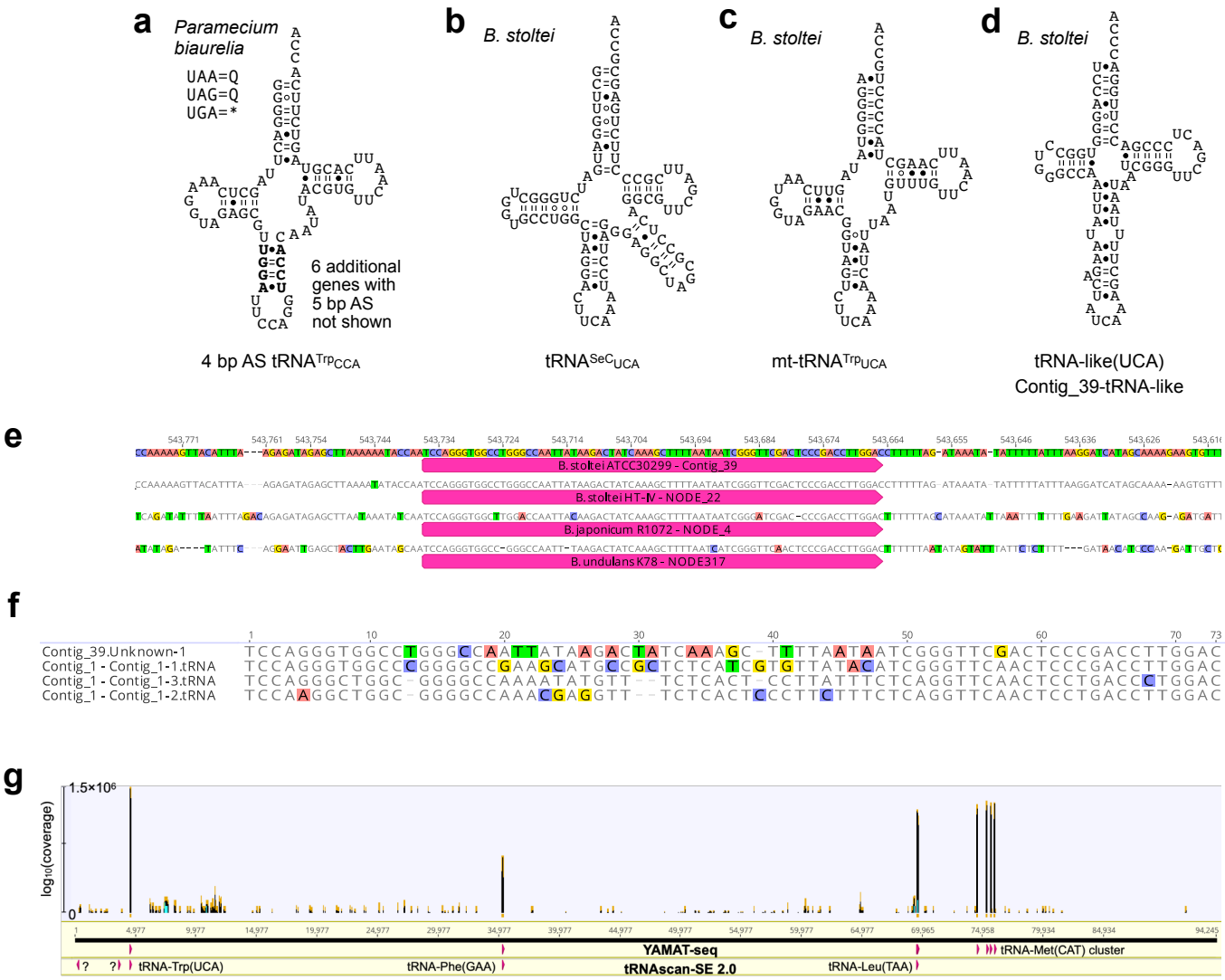
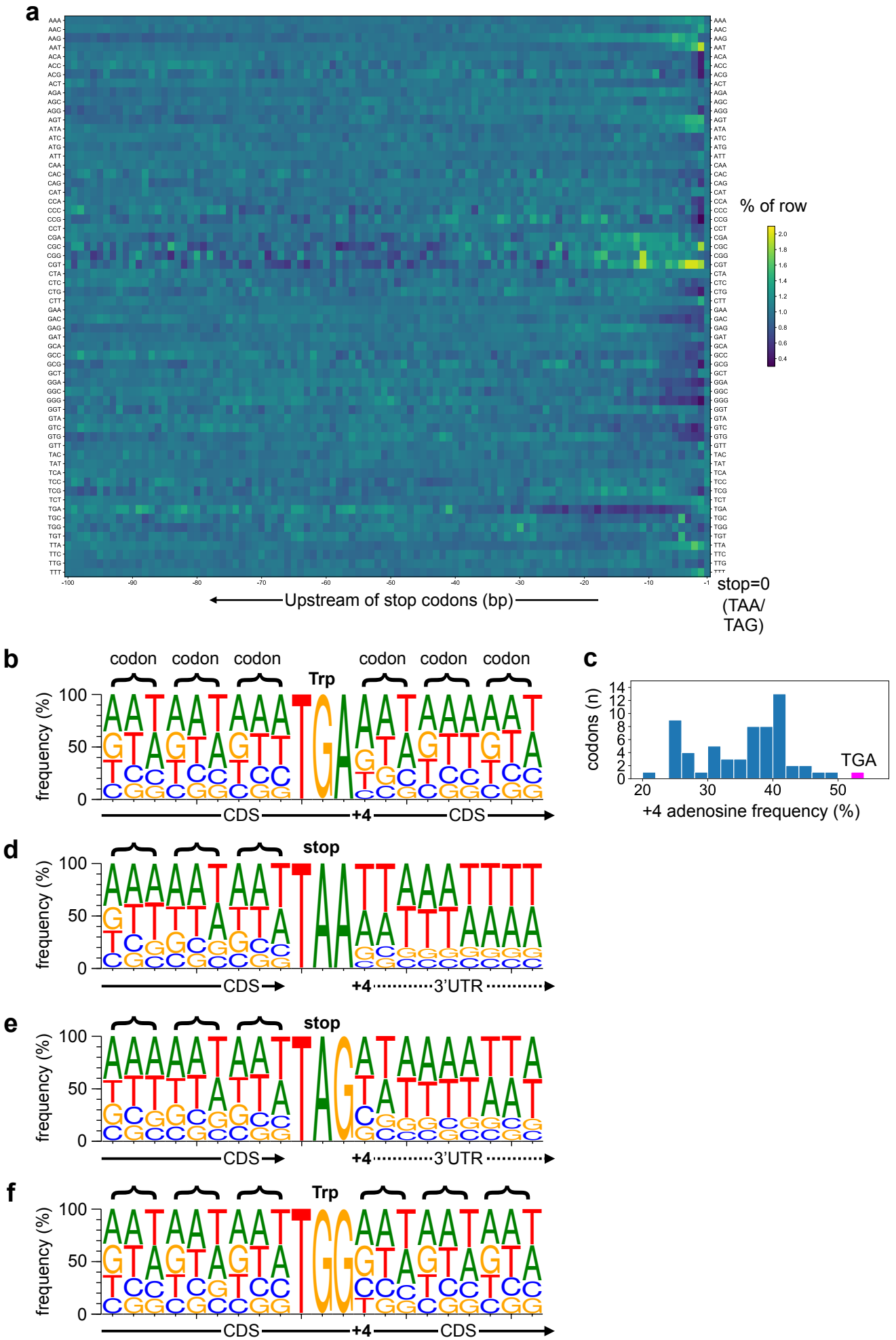

Condylostoma magnum
UAA=Q/*
UAG=Q/*
UGA=W/*
8 genes
No C to U editing
Blepharisma stoltei
UAA=*
UAG=*
UGA=W
6 genes
No C to U editing
Stentor coeruleus
UAA=*
UAG=*
UGA=*
5 genes
Loxodes magnus
UAA=Q/*?
UAG=Q
UGA=W/*
102 genes
No C to U editing
1 additional putative pseudogene not shown
Loxodes magnus
3 putative pseudogenes
deletion
3 additional putative pseudogenes
Fabrea salina
UAA=*
UAG=*
UGA=*
2 genes

Fig 2


a
eRF1 window
57
67
Genetic code:
UAA
UAG
UGA
UGA=✶ + Ser67
UGA= W, W/✶ + Ala/Gly67
Saccharomyces cerevisiae
Amoebophyra sp. ex K. veneficum
Giardia intestinalis
Blastocrithida nonstop
Trypanosoma brucei
Plagiopyla frontata
Tetrahymena thermophila
Tetrahymena thermophila
Pseudocohnilembus persalinus
Nyctotherus ovalis
Nyctotherus ovalis
Euplotes octocarinatus
Euplotes octocarinatus
Euplotes octocarinatus
Geleia acuta
Parduczia sp.
Kentrophoros sp.
Trachelocercidae sp.
Remanella sp.
Loxodes magnus
Loxodes magnus
Geleia acuta
Spirostomum ambiguum
Condylostoma magnum
Climacostomum virens
Fabrea salina
Stentor coeruleus
Stentor coeruleus
Stentor coeruleus
Blepharisma stoltei
Blepharisma stoltei
Blepharisma stoltei
Ciliophora
Karyorelictea
Heterotrichea


b
TGA
ATG
AGT
GAT
GTA
% of row
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
-100
-90
-80
-70
-60
-50
-40
-30
-20
-10
-1
stop=0
(TAA/TAG)
Upstream of stop codons (bp)

Extended Data Fig 1



Extended Data Fig 2