

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

## Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

[About eLife's process](#)

## Reviewed preprint version 1

February 19, 2024 (this version)

## Sent for peer review

December 15, 2023

## Posted to preprint server

October 11, 2023

Estienne Carl Swart , Christiane Emmerich, Kwee Boon Brandon Seah, Minakshi Singh, Yekaterina Shulgina, Aditi Singh

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

 [https://en.wikipedia.org/wiki/Open\\_access](https://en.wikipedia.org/wiki/Open_access)

 Copyright information

## Abstract

Ciliates are a widespread clade of microbial eukaryotes with the greatest diversity of nuclear genetic codes (at least eight) following a recent addition<sup>1</sup>. All non-standard ciliate genetic codes involve stop codon reassignments<sup>1,2,3</sup>. Two of these codes are ambiguous<sup>1–3</sup>, with “stop” codons either translated or terminating translation depending on their context<sup>2,3</sup>. Ambiguous genetic codes have arisen not only in ciliates, but also independently in trypanosomatids from the genus *Blastocrithidia*<sup>4</sup> and an alveolate species from the genus *Amoebophrya*<sup>5</sup>. 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<sup>1,2</sup>. Furthermore, though a UGA-cognate tRNA<sup>Cys</sup><sub>UCA</sub> was reported in *Euplotes*<sup>6</sup>, a ciliate genus that translates UGA as cysteine, vexingly, no nuclear genome-encoded tRNA<sup>Trp</sup><sub>UCA</sub> 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 *Condyllostoma magnum* using 4 base pair anticodon stem (AS) near-cognate tryptophan tRNA<sup>Trp</sup><sub>CCA</sub>’s, rather than the typical 5 base pair stem tRNAs<sup>7</sup>. 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<sup>7</sup>. 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.

## eLife assessment

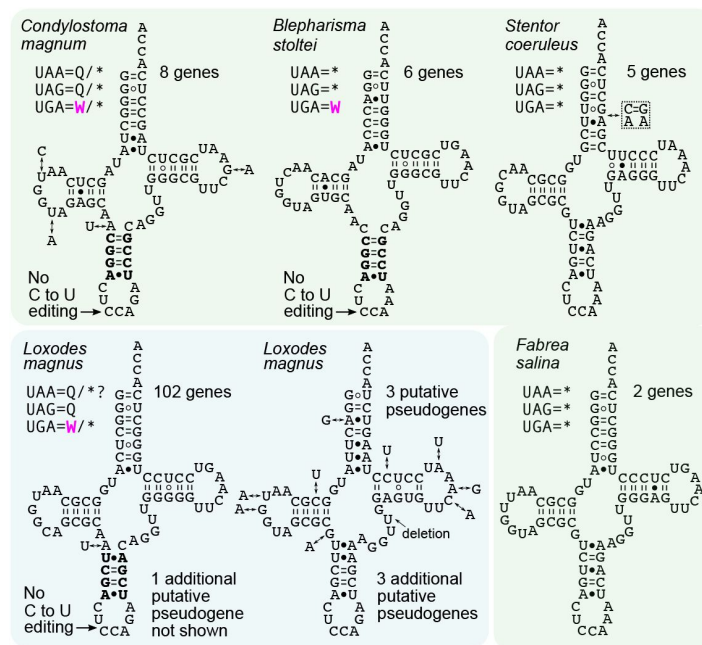
The manuscript explores the ways in which the genetic code evolves, specifically how stop codons are reassigned to become sense codons. The authors present phylogenetic data showing that mutations at position 67 of the termination factor are present in organisms that nevertheless use the UGA codon as a stop codon, thereby questioning the importance of this position in the reassignment of stop codons. Alternative models on the role of eRF1 would reflect a more balanced view of the data. Overall, the data are **solid** and these findings will be **valuable** to the genomic/evolution fields.

## 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<sup>7</sup>, likely originates from a bacterium present in the ciliate culture, not from *Condyllostoma magnum*. Thus, *Condyllostoma* probably only has nuclear genome-encoded 4 bp AS tRNA<sup>Trp</sup><sub>CCA</sub>'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 *Condyllostoma magnum* (class Heterotrichea; UAR=Q/\*, UGA=W/\*)<sup>2,3</sup> 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/\*)<sup>8</sup>, and in the ciliate genus *Blepharisma* (class Heterotrichea; **Fig. 1**), which appears to translate UGA unambiguously as tryptophan (UAR=\*, UGA=W)<sup>8</sup>. A draft *Loxodes magnus* somatic genome assembly has over one hundred 4 bp AS tRNA<sup>Trp</sup> genes but just six 5 bp AS tRNA<sup>Trp</sup> 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<sup>9,10</sup>, *Stentor coeruleus* and *Fabrea salina*, only have 5 bp AS tRNA<sup>Trp</sup> genes (**Fig. 1**). Among members of the *Paramecium aurelia* complex, which are not known to translate UGA as tryptophan (code: UAR=Q, UGA=\*)<sup>2</sup>, we found a ciliate species, *Paramecium biaurelia*, with at least one 4 bp AS tRNA (**Extended Data Fig. 1a**) among its eight tRNA<sup>Trp</sup> 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<sup>Trp</sup><sub>CCA</sub>'s generates UCA anticodons that enable mitochondrial UGA codon translation<sup>11</sup>. Kachale et al. reported no editing of cytosolic *Blastocrithidia* tRNA<sup>Trp</sup><sub>CCA</sub>'s<sup>7</sup> that could permit translation of cytosolic UGAs. Previously, tRNA sequencing did not reveal C-to-U editing of the *Condyllostoma magnum* tRNA<sup>Trp</sup><sub>CCA</sub> anticodon wobble base that would generate an anticodon complementary to UGA



**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<sub>UCA</sub> predictions (selenocysteine and mitochondrial). *Blepharisma japonicum* and *Blepharisma undulans* tRNA<sup>Trp</sup><sub>CCA</sub> also have 4 bp anticodon stems (Source Data Fig. 1 and **Extended Data Fig. 1**).

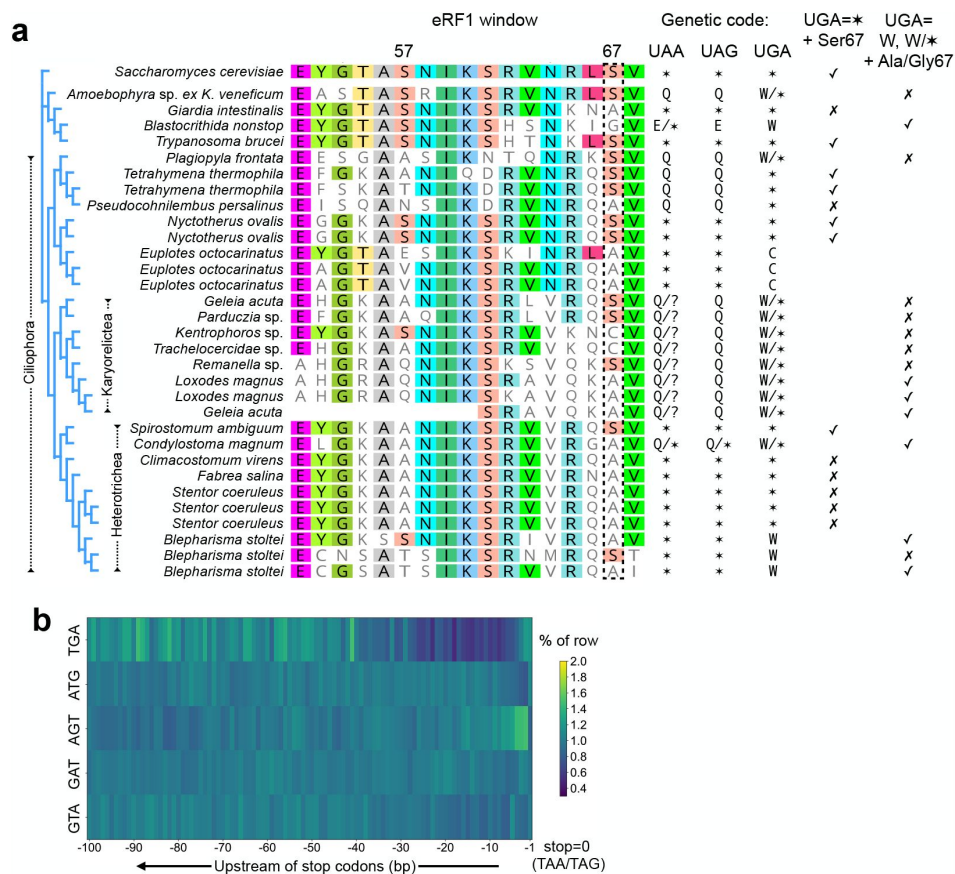
codons<sup>2</sup>. We have also observed no appreciable C-to-U editing (> 0.1% of tryptophan tRNAs) in *Blepharisma* and *Loxodes* 4 bp AS tRNA<sup>Trp</sup> sequences (in 172,929 and 9,721 unique reads, respectively; Supplementary Table 1, Source Data Fig. 1).

Anticipating the discovery of natural 4 bp AS tRNA<sup>Trp</sup><sub>CCA</sub>'s in *Blastocrithidia* and the ciliates *Condyllostoma*, *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<sup>12,13</sup>. 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<sup>12</sup>, which involves G-U wobble pairing at the 1<sup>st</sup> codon position that is normally disallowed. In the second paper, using su7 tRNA<sub>CUA</sub> they showed that mutations that disrupt the top AS stem base pair promote UAA translation<sup>13</sup>, which involves mismatched 3<sup>rd</sup> position A-C like that required for UGA translation by 4bp AS tRNA<sub>CCA</sub>'s<sup>7</sup>.

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<sup>14,15</sup>. With the benefit of additional eRF1 sequences and ciliate genetic codes, we previously reported multiple counterexamples to such associations<sup>2</sup>. 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<sup>Trp</sup> anticodon stem for efficient UGA translation as tryptophan<sup>7</sup>. 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 *Condyllostoma* 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)<sup>8</sup>. 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/\*)<sup>1</sup>. 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*<sup>16,17,18</sup>. 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<sup>19</sup>) 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”<sup>16</sup>. It is conceivable that these paralogs have functionally diverged, now serving an alternative role like Dom34/Pelota.



**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**.

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<sup>20</sup>. Correspondingly, with some capacity of *Blepharisma* eRF1 to recognize UGA as stop codons, a signal of potential competition between *B. stoltei* eRF1 and tRNA<sup>Trp</sup> 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<sup>2,8</sup>. 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<sup>7</sup>. 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”<sup>21</sup>. 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.

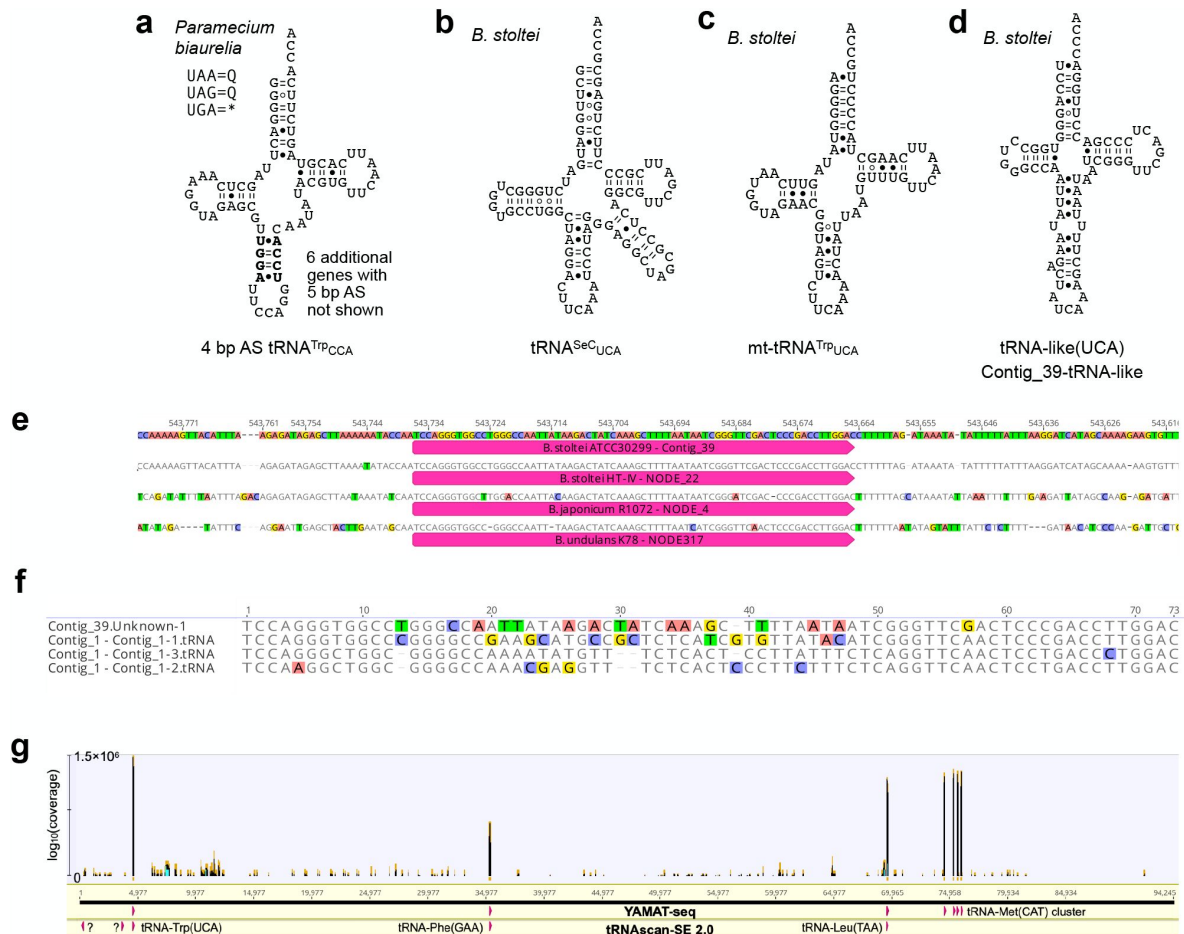
## Contributions

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.

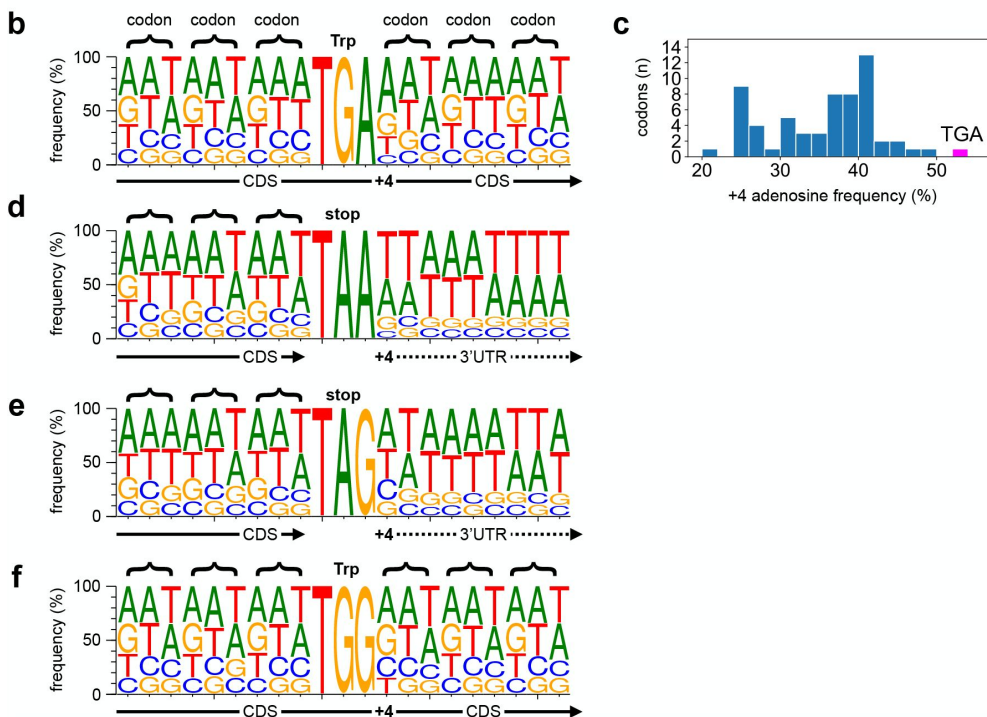
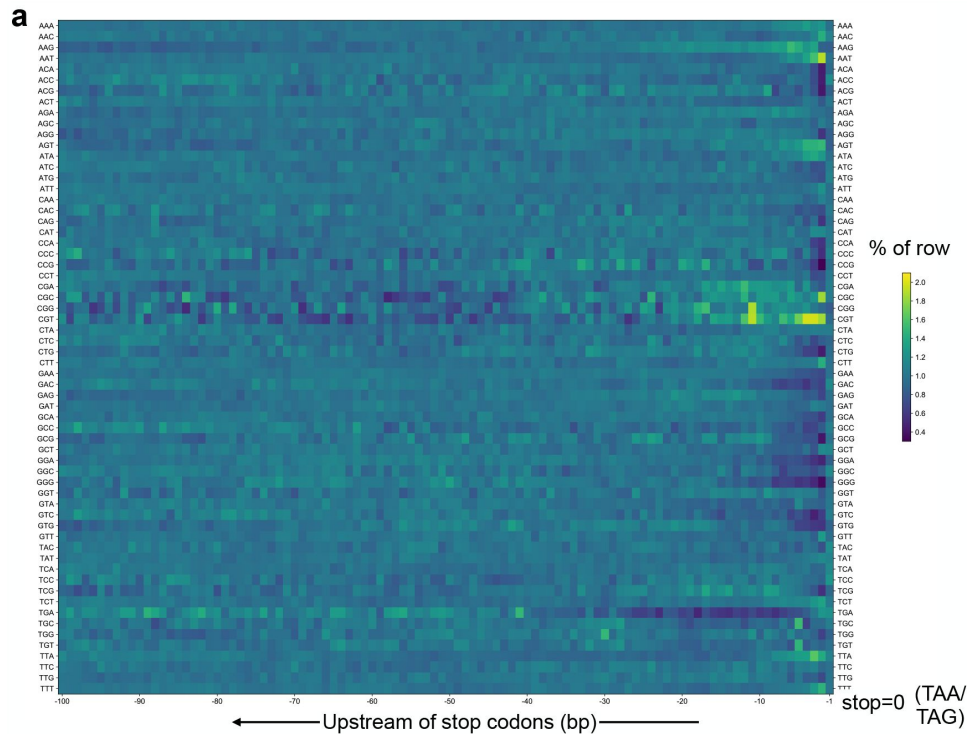
## Figure captions



**Extended Data Fig. 1.**

### Secondary structures of *Parametium biaurelia* 4 bp AS tRNA<sup>Trp</sup><sub>CCA</sub> and *Blepharisma* tRNA<sub>UCA</sub>'s and a tRNA-like molecule with a possible UCA anticodon.

(a) *Parametium biaurelia* tRNA<sup>Trp</sup><sub>CCA</sub>; terminal nucleotides predicted by tRNAscan-SE 2.0. (b) *B. stoltei* tRNA<sup>Sec</sup><sub>UCA</sub>. (c) *B. stoltei* mitochondrial tRNA<sup>Trp</sup><sub>UCA</sub>. (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).

## References

1. McGowan J., et al. (2022) **Identification of a Non-Canonical Ciliate Nuclear Genetic Code Where UAA and UAG Code for Different Amino Acids** *BioRxiv* <https://doi.org/10.1101/2022.12.16.520718>
2. Swart E. C., Serra V., Petroni G., Nowacki M. (2016) **Genetic codes with no dedicated stop codon: Context-dependent translation termination** *Cell* **166**:691–702
3. Heaphy S. M., Mariotti M., Gladyshev V. N., Atkins J. F., Baranov P. V. (2016) **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
4. Záhonová K., Kostygov A. Y., Ševčíková T., Yurchenko V., Eliáš M. (2016) **An Unprecedented Non-canonical Nuclear Genetic Code with All Three Termination Codons Reassigned as Sense Codons** *Curr. Biol* **26**:2364–2369
5. Bachvaroff T. R. (2019) **A precededented nuclear genetic code with all three termination codons reassigned as sense codons in the syndinean *Amoebophrya* sp. ex *Karlodinium veneficum*** *PLoS ONE* **14**
6. Turanov A. A., et al. (2009) **Genetic code supports targeted insertion of two amino acids by one codon** *Science* **323**:259–261
7. Kachale A., et al. (2023) **Short tRNA anticodon stem and mutant eRF1 allow stop codon reassignment** *Nature* **613**:751–758
8. Seah B. K. B., Singh A., Swart E. C. (2022) **Karyorelict ciliates use an ambiguous genetic code with context-dependent stop/sense codons** *Peer Community Journal* **2**
9. Slabodnick M. M., et al. (2017) **The macronuclear genome of *Stentor coeruleus* reveals tiny introns in a giant cell** *Curr. Biol* **27**:569–575
10. Zhang B., et al. (2022) **An Extremely Streamlined Macronuclear Genome in the Free-Living Protozoan *Fabrea salina*** *Mol. Biol. Evol* **39**
11. Alfonzo J. D., Blanc V., Estévez A. M., Rubio M. A., Simpson L. (1999) **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
12. Schultz D. W., Yarus M. (1994) **tRNA structure and ribosomal function. I. tRNA nucleotide 27-43 mutations enhance first position wobble** *J. Mol. Biol* **235**:1381–1394
13. Schultz D. W., Yarus M. (1994) **tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations** *J. Mol. Biol* **235**:1395–1405
14. Lehman N. (2001) **Molecular evolution: Please release me, genetic code** *Curr. Biol* **11**:R63–6
15. 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

16. Atkinson G. C., Baldauf S. L., Hauryliuk V. (2008) **Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components** *BMC Evol. Biol* **8**
17. Inagaki Y., Doolittle W. F. (2001) **Class I release factors in ciliates with variant genetic codes** *Nucleic Acids Res* **29**:921–927
18. Liang H., Wong J. Y., Bao Q., Cavalcanti A. R. O., Landweber L. F. (2005) **Decoding the decoding region: analysis of eukaryotic release factor (eRF1) stop codon-binding residues** *J. Mol. Evol* **60**:337–344
19. Swart E. (2022) **Blepharisma stoltei ATCC 30299 Macronuclear genome** *Edmond* <https://doi.org/10.17617/3.8c>
20. Eliseev B. D., et al. (2011) **Translation termination factor eRF1 of the ciliate Blepharisma japonicum recognizes all three stop codons** *Mol Biol (NY)* **45**:614–618
21. Schultz D. W., Yarus M. (1996) **On malleability in the genetic code** *J. Mol. Evol* **42**:597–601

## Article and author information

### Estienne Carl Swart

Max Planck Institute for Biology, Tuebingen, Germany

**For correspondence:** [estienne.swart@tuebingen.mpg.de](mailto:estienne.swart@tuebingen.mpg.de)

ORCID iD: [0000-0002-1362-1602](https://orcid.org/0000-0002-1362-1602)

### Christiane Emmerich

Max Planck Institute for Biology, Tuebingen, Germany

### Kwee Boon Brandon Seah

Max Planck Institute for Biology, Tuebingen, Germany

ORCID iD: [0000-0002-1878-4363](https://orcid.org/0000-0002-1878-4363)

### Minakshi Singh

Max Planck Institute for Biology, Tuebingen, Germany

ORCID iD: [0000-0001-9674-6162](https://orcid.org/0000-0001-9674-6162)

### Yekaterina Shulgina

Department of Molecular and Cellular Biology, Harvard University, Cambridge, United States

ORCID iD: [0000-0001-7658-9294](https://orcid.org/0000-0001-7658-9294)

### Aditi Singh

Max Planck Institute for Biology, Tuebingen, Germany

ORCID iD: [0000-0001-7840-9058](https://orcid.org/0000-0001-7840-9058)

## Copyright

© 2024, Swart et al.

This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Editors

Reviewing Editor

**Yamini Dalal**

National Cancer Institute, Bethesda, United States of America

Senior Editor

**Yamini Dalal**

National Cancer Institute, Bethesda, United States of America

### Reviewer #1 (Public Review):

The issue:

The ciliates are a zoo of genetic codes, where there have been many reassignments of stop codons, sometimes with conditional meanings which include retention of termination function, and thus > 1 meaning. Thus ciliate coding provides a hotspot for the study of genetic code reassignments.

The particular issue here is the suggestion that translation of a stop (UGA) in *Blastocytisidia* has been attributed to a joint change in the protein release factor that reads UGA's and also breaking a base pair at the top of the anticodon stem of tRNA<sup>Trp</sup> (Nature 613, 751, 2023).

The work:

However, Swart, et al have looked into this suggestion, and find that the recently suggested mechanism is overly complicated.

The broken pairing at the top of the anticodon stem of tRNA<sup>Trp</sup> indeed accompanies the reading of UGA as Trp as previously suggested. It changes the codon translated even though the anticodon remains CCA, complementary to UGG. A compelling point is that this misreading matches previous mutational studies of *E. coli* tRNA's, in which breaking the same base pair in a mutant tRNA<sup>Trp</sup> suppressor tRNA stimulated the same kind of miscoding.

But the amino acid change in release factor eRF1, the protein that catalyzes termination of protein biosynthesis at UGA is broadly distributed. There are about 9 organisms where this mutation can be compared with the meaning of UGA, and the changes are not highly correlated with a change in the meaning of the codon. Therefore, because UGA can be translated as Trp with or without the eRF1 mutation, Swart et al suggest that the tRNA anticodon stem change is the principal cause of the coding change.

The review:

Swart et al have a good argument. I would only add that eRF1 participation is not ruled out, because finding that UGA encodes Trp does not distinguish between encoding Trp 90% of the time and encoding it 99% of the time. The release factor could still play a measurable quantitative role, but the major inference here seems convincing.

<https://doi.org/10.7554/eLife.93502.1.sa1>

### Reviewer #2 (Public Review):

The manuscript raises interesting observations about the potential evolution of release factors and tRNA to readdress the meaning of stop codons. The manuscript is divided into two parts: The first consists of revealing that the presence of a trp tRNA with an AS of 5bp in *Condyllostoma magnum* is probably linked to contamination in the databases by sequences from bacteria. This is an interesting point which seems to be well supported by the data

provided. It highlights the difficulty of identifying active tRNA genes from poorly annotated or incompletely assembled genomes. The second part criticises the fact that a mutation at position S67 of eRF1 is required to allow the UGA codon to be reassigned as a sense codon. As supporting evidence, they provide a phylogenetic study of the eRF1 factor showing that there are numerous ciliates in which this position is mutated, whereas the organism shows no trace of the reassignment of the UGA codon into a sense codon. While this criticism seems valid at first glance, it suffers from the lack of information on the level of translation of UGA codons in the organisms considered. It has been clearly shown that S67G or S67A mutations allow a strong increase in the reading of UGA codons by tRNAs, so this point is not in doubt. However, this has been demonstrated in model organisms, and we now need to determine whether other changes in the translational apparatus could accompany this mutation by modifying its impact on the UGA codon. This is a point partly raised at the end of the manuscript. Indeed, it is quite possible that in these organisms the UGA codon is both used to complete translation and is subject to a high level of readthrough. Actually, in the presence of a mutation at position 67 (or elsewhere), the reading of the UGA can be tolerated under specific stress conditions (nutrient deficiency, oxidative stress, etc.), so the presence of this mutation could allow translational control of the expression of certain genes. On the other hand, it seems obvious to me that there are other ways of reading through a stop codon without mutating eRF1 at position S67. So the absence of a mutation at this position is not really indicative of a level of reading of the UGA codon. Before writing such a strong assertion as that found on page 3, experiments should be carried out. The authors should therefore moderate their assertion.

To make a definitive conclusion, we would need to be able to measure the level of termination and readthrough in these organisms. So, from my point of view, all the arguments seem rather weak. Moreover, the authors themselves indicate that the conjunction between a Trp tRNA that is efficient at reading the UGA codon and an eRF1 factor that is not efficient at recognising this stop codon could be the key to reassignment.

<https://doi.org/10.7554/eLife.93502.1.sa0>

## Author Response

### **eLife assessment**

*The manuscript explores the ways in which the genetic code evolves, specifically how stop codons are reassigned to become sense codons. The authors present phylogenetic data showing that mutations at position 67 of the termination factor are present in organisms that nevertheless use the UGA codon as a stop codon, thereby questioning the importance of this position in the reassignment of stop codons. Alternative models on the role of eRF1 would reflect a more balanced view of the data. Overall, the data are solid and these findings will be valuable to the genomic/evolution fields.*

### **Public Reviews:**

#### **Reviewer #1 (Public Review):**

*The issue:*

*The ciliates are a zoo of genetic codes, where there have been many reassignments of stop codons, sometimes with conditional meanings which include retention of termination function, and thus > 1 meaning. Thus ciliate coding provides a hotspot for the study of genetic code reassignments.*

*The particular issue here is the suggestion that translation of a stop (UGA) in *Blastocytisidia* has been attributed to a joint change in the protein release factor that reads UGA's and also breaking a base pair at the top of the anticodon stem of tRNA<sup>Trp</sup> (Nature 613, 751, 2023).*

*The work:*

*However, Swart, et al have looked into this suggestion, and find that the recently suggested mechanism is overly complicated.*

*The broken pairing at the top of the anticodon stem of tRNA<sup>Trp</sup> indeed accompanies the reading of UGA as Trp as previously suggested. It changes the codon translated even though the anticodon remains CCA, complementary to UGG. A compelling point is that this misreading matches previous mutational studies of *E coli* tRNA's, in which breaking the same base pair in a mutant tRNA<sup>Trp</sup> suppressor tRNA stimulated the same kind of miscoding.*

This is a fair characterization, and we would also note the additional positive aspect: that we observed there is consistency in the presence of 4 bp tRNA-Trp anticodon stems in those ciliates which translate UGA as tryptophan, and generally 5 bp anticodon stems in those that do not (including *Euplotes* with UGA=Cys).

*But the amino acid change in release factor eRF1, the protein that catalyzes termination of protein biosynthesis at UGA is broadly distributed. There are about 9 organisms where this mutation can be compared with the meaning of UGA, and the changes are not highly correlated with a change in the meaning of the codon. Therefore, because UGA can be translated as Trp with or without the eRF1 mutation, Swart et al suggest that the tRNA anticodon stem change is the principal cause of the coding change.*

We do think multiple lines of evidence support the shorter tRNA anticodon stem promoting UGA translation, but also think other changes in the translation system may be important. For instance, structural studies suggest interaction of ribosomal RNA with extended stop codons (particularly the base downstream of the triplet) during translation termination (Brown et al. 2015, Nature). As we noted, previous studies have sought to correlate individual eRF1 substitutions with genetic code changes, but the proposed correlations have invariably disappeared once new tranches of eRF1 sequences and alternative genetic codes for different species became available. This is why we concluded that there needs to be more focus on obtaining and understanding molecular structures during translation termination, particularly in the organisms with alternative codes.

*The review:*

*Swart et al have a good argument. I would only add that eRF1 participation is not ruled out, because finding that UGA encodes Trp does not distinguish between encoding Trp 90% of the time and encoding it 99% of the time. The release factor could still play a measurable quantitative role, but the major inference here seems convincing.*

We agree that eRF1 may participate and compete with the tRNA, but we question the hypothesis that the particular amino acid position/substitution proposed by Kachale et al. 2023 is the key. There is experimental evidence in the form of Ribo-seq for the ciliate *Condylostoma magnum* (A67), which does appear to efficiently translate UGA sense codons (Swart et al. 2016, Figure S3: <https://doi.org/10.1016/j.cell.2016.06.020>): we observed no dip in ribosome footprints downstream of these codons, as there would be in the case of classical translational readthrough in standard genetic code organisms (which is usually relatively inefficient - certainly well below 50% of upstream translation from our reading of the

literature). Ribo-seq also supports efficient termination at those *Condylostoma* UGA codons that are stops.

Of course, the entire translation system may have evolved to be as efficient as what we currently observe, and it is not unreasonable to consider that it may have been less efficient in the past. However, not so inefficient that the error rate incurred would have been strongly deleterious. Importantly also, we believe the role of multiple eRF1 paralogs in translation termination in the ciliates really needs to be investigated, given that translation is inherently probabilistic with any of these proteins potentially being incorporated into the ribosome.

**Reviewer #2 (Public Review):**

*The manuscript raises interesting observations about the potential evolution of release factors and tRNA to readdress the meaning of stop codons. The manuscript is divided into two parts: The first consists of revealing that the presence of a trp tRNA with an AS of 5bp in *Condylostoma magnum* is probably linked to contamination in the databases by sequences from bacteria. This is an interesting point which seems to be well supported by the data provided. It highlights the difficulty of identifying active tRNA genes from poorly annotated or incompletely assembled genomes.*

We will consider adding subheadings in revising the manuscript to make the structure more explicit, as it really has three parts to it, with the third largely in the supplement. The “good” was that there is a range of support for the 4 bp AS stem, with new evidence we supplied from ciliates and older studies with *E. coli* tRNAs. The “bad” is that scrutiny of eRF1 sequences, with the addition of ones we provided, contradicts the hypothesis by Kachale et al. that a S67A/G substitution is necessary for genetic code evolution in *Blastocrithidia* and certain ciliates. The “ugly” is that a tRNA shown in a main figure in Kachale et al. 2023, and which was investigated in a number of subsequent experiments, is almost certainly a bacterial contaminant.

Proper scrutiny of the bacterial tRNA should have led to its immediate recognition and rejection, as one of us did years ago in searches of tRNAs in a preliminary *Condylostoma* genome assembly (only predicted 4 bp AS tRNA secondary structures were shown in Swart et al. 2016, Fig S4B and C). Evidence for the bacterial nature of this tRNA was placed in the supplement of the present manuscript, as the meat of the critique was the consideration of the evidence for and against its good and bad aspects. The bacterial tRNA secondary structure has been removed from the main figure by Kachale et al. 2023, and downstream experiments based on synthetic constructs for this tRNA have also been revised (<https://www.nature.com/articles/s41586-024-07065-0>).

Much of the rest of the supplement served to correct multiple errors in genetic codes in public sequence databases that led to additional errors and difficulties in interpreting the eRF1 substitutions in Kachale et al. 2023. It is important that these codes get corrected. If not they create multiple headaches for users besides those investigating genetic codes, as we found out in communications with authors and a colleague of Kachale et al. 2023 (in particular, leading to thousands of missing genes in the macronuclear genome of the standard code ciliate *Stentor coeruleus* that were removed in automated GenBank processing due to incorrectly having an alternative genetic code specified).

Recently the NCBI Genetic Codes curators reinstated a genetic code incorrectly attributed to the ciliate *Blepharisma* (“*Blepharisma* nuclear genetic code”) (<https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi#SG15>), despite us requesting a reasonable fix years ago. This would be very confusing for those that are not in the know. We have explained this confusion in our supplement too. Thus we also hope that this paper will aid in communication with the genetic code database curators and in correcting such issues.

*The second part criticises the fact that a mutation at position S67 of eRF1 is required to allow the UGA codon to be reassigned as a sense codon. As supporting evidence, they provide a phylogenetic study of the eRF1 factor showing that there are numerous ciliates in which this position is mutated, whereas the organism shows no trace of the reassignment of the UGA codon into a sense codon. While this criticism seems valid at first glance, it suffers from the lack of information on the level of translation of UGA codons in the organisms considered.*

Firstly, we not only showed that there are organisms with the S67 substitution but no UGA reassignment, but also provided evidence for the converse: organisms with a UGA=Trp reassignment but without the S67 substitution (both ciliates and a non-ciliate). So, two related lines of substitutions were not consistent with the eRF1 substitution hypothesis proposed.

Secondly, we disagree that there is a “lack of information about UGA translation in the organisms considered”. Evolution has already supplied information as to whether UGA codons are translated at an appreciable level in the organisms of interest, in the form of codon frequencies within their protein-coding sequences and those ending them. If UGA was translated at appreciable levels, it would be found at a corresponding frequency in coding sequences. In genomes with thousands of genes, if not predicted as amino acids, they likely primarily serve as stops. Low levels of potential readthrough of actual stops would not change the arguments. With the exception of selenocysteine translation (which is restricted to a limited number of genes by the condition of requiring a specific mRNA secondary structure) there is no expectation of meaningful levels of UGA translation when this codon is missing from the bulk of coding sequences (CDSs).

This is well illustrated by the heterotrichs, a clade of ciliates that use a variety of genetic codes. In heterotrichs that use the standard code, UGA is virtually absent from coding sequences, only appearing at the 3' end of transcripts in the predicted stop codon and 3'-UTR (Seah et al. 2022, Figure 5). This contrasts notably with other genera like *Blepharisma* where appreciable levels of UGA codons occur throughout coding sequences, upstream of the predicted UAA and UAG stops (Seah et al. 2022, Figure 5: <https://www.biorxiv.org/content/biorxiv/early/2022/07/12/2022.04.12.488043/F5.large.jpg>). The difference in the UGA, UAG and UAA codon frequencies in 3' UTRs compared to the upstream frequencies in CDSs of standard genetic code heterotrichs is stark. Frequencies of all three codons are elevated in the 3' UTRs of all heterotrich ciliates, irrespective of their genetic codes (Seah et al. 2022, Figure 5), according with these codons not being deleterious in this region and strongly selected against upstream, within CDSs.

The reviewer raises the possibility that UGA may appear to be a stop codon but still have biologically significant translational readthrough. We think that this is unlikely in the heterotrich ciliate species discussed here, which have extremely short (median 21-26 bp) and AU-rich 3'-UTRs compared to yeast and animals (Seah et al. 2022). Therefore, in heterotrichs where UGA is predicted to be a stop, translational readthrough would lead to extensions of only a few amino acids and be relatively inconsequential, as there are plenty of secondary UAA, UAG and UGA codons downstream of the typical stop.

If one were to consistently pursue the reviewer's line of argumentation, one would also have to argue against the very reasoning used in Kachale et al. 2023 about all the stop codon predictions/reassignments in protists for which experiments were not conducted in *S. cerevisiae* or other translation systems, as well as decades of prior work using sequence conservation in multiple sequence alignments to infer alternative genetic codes.

Furthermore, experimental information for UGA translation levels is available for the ciliate *Condylostoma magnum*, predominantly in the form of Ribo-seq (Swart et al. 2016). Similarly to *Condylostoma*'s UAA and UAG codons, Ribo-seq shows that the UGA codons are generally

either efficiently translated when present in the bodies of CDSs or terminate translation as actual stops close to mRNA 3' termini/poly(A) tails (Swart et al. 2016). Thus, irrespective of the presence of the hypothesized eRF1 substitution there is an example of relatively discrete reading of UGA codons in ciliates as either stops or amino acids. This contrasts with Kachale et al 2023's experiments in yeast with yeast eRF1 S67G or Blastocrithida eRF1 which also has glycine at the equivalent position that appear to lead to modest readthrough. In addition, efficient reading of codons in either of two ways also occurs in the ciliate genus Euplotes in which "stop" codons can either serve as frameshift sites during translation within coding sequences or be actual stops when they are close to 3' mRNA termini (Lobanov et al. 2017), as verified by Ribo-seq and protein mass spectrometry.

*It has been clearly shown that S67G or S67A mutations allow a strong increase in the reading of UGA codons by tRNAs, so this point is not in doubt. However, this has been demonstrated in model organisms, and we now need to determine whether other changes in the translational apparatus could accompany this mutation by modifying its impact on the UGA codon. This is a point partly raised at the end of the manuscript.*

There is no doubt that S67G or S67A mutations lead to increased translational readthrough, but this is restricted to experiments with or in baker's yeast or other standard genetic code surrogate model organisms. Experiments introducing eRF1 sequences from alternative genetic code eukaryotes into translation systems of such standard genetic code eukaryotes are not compelling because the rest of the associated translation system has also evolved tremendously. As far as we are aware, no in vivo experiments with ciliate eRF1s have been conducted to determine if position 67 or other substitutions have any effect. These considerations are critical given the vast evolutionary distances between yeasts, Blastocrithidia, the ciliates and Amoeboophrya sp. ex Karlodinium veneficum. On the other hand, the evolutionary information presented contradicts the importance of this substitution in the Amoeboophrya species and ciliates. We will consider how to incorporate these ideas in the revised version of the manuscript.

*Indeed, it is quite possible that in these organisms the UGA codon is both used to complete translation and is subject to a high level of readthrough. Actually, in the presence of a mutation at position 67 (or elsewhere), the reading of the UGA can be tolerated under specific stress conditions (nutrient deficiency, oxidative stress, etc.), so the presence of this mutation could allow translational control of the expression of certain genes.*

As explained a couple replies above, it is not constructive to invoke the additional complexity of conditional translation or any other kinds of factors that lead to enhanced readthrough, because the translation of UGA sense codons in the ciliate Condyllostoma, where we have supporting experimental evidence, does not resemble translational readthrough. These codons occur in constitutively expressed single-copy genes, like a tryptophan tRNA synthetase and an eRF1 protein (Swart et al. 2016), not ones that might be expected to be conditionally translated.

*On the other hand, it seems obvious to me that there are other ways of reading through a stop codon without mutating eRF1 at position S67. So the absence of a mutation at this position is not really indicative of a level of reading of the UGA codon.*

It may seem obvious to the reviewer, but that is neither what Kachale et al. originally proposed nor what we questioned. Kachale et al. hypothesized that mutation of S67 to A or G is necessary for UGA=Trp translation, but we provided evidence that it is not: multiple organisms with S67 or C67 that translate UGA as tryptophan. Kachale et al. also originally suggested that the S67 to A/G substitution is also necessary in Condyllostoma for UGA

translation as tryptophan by weakening its recognition of this codon as a stop (from their abstract: “Virtually the same strategy has been adopted by the ciliate *Condylostoma magnum*.”). However, as we have stated, *Condylostoma* (A67) is both able to efficiently terminate at UGA stop codons and to efficiently translate (other) UGA sense codons, which does not fit this hypothesis.

*Before writing such a strong assertion as that found on page 3, experiments should be carried out. The authors should therefore moderate their assertion.*

Experiments should be carried out in the organisms in which stop codon reassignments have readily occurred and their close relatives that have not, not distantly related ones where they rarely, if ever, occur, like yeasts. We made this point in the conclusion. There is too much emphasis on models for investigation of genetic code evolution via stop codon reassignments in questionable models and too little investigation in the really good ones, particularly the ciliates. This clade has genera that are amenable to molecular experiments including *Paramecium*, *Tetrahymena* and *Oxytricha*. We plan to add some text about these considerations in revision.

*To make a definitive conclusion, we would need to be able to measure the level of termination and readthrough in these organisms. So, from my point of view, all the arguments seem rather weak.*

We reiterate: there is experimental information about translation and termination in two ciliate species worth considering, including one that translates UGA codons depending on their context. If one chooses to ignore the evolutionary information presented, this not only ignores all prior approaches to infer genetic codes, but also the fact that there is experimental verification and other lines of evidence supporting these approaches.

*Moreover, the authors themselves indicate that the conjunction between a Trp tRNA that is efficient at reading the UGA codon and an eRF1 factor that is not efficient at recognising this stop codon could be the key to reassignment.*

This does not convey well what we wrote, since the main consideration was overall eRF1 structure, rather than individual amino acid substitutions. Here are the key sentences:

“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.”