MAX PLANCK INSTITUTE
FOR BIOLOGY TÜBINGEN

MPI for Biology Tübingen, Max-Planck-Ring 9, 72076 Tübingen, Germany

Research Group
Ciliate Biology

Max-Planck-Ring 5
D-72076 Tübingen

Estienne Swart
Phone: +49 (0)7071-601-340
estienne.swart@tuebingen.mpg.de

For the attention of:
Assoc. Prof Tomas Dolezal, head of the PhD program board,
University of South Bohemia
Ceske Budejovice, 37005
Czech Republic

Review of the doctoral thesis of Ambar Kachale

Overall comments:

It was a pleasure to read this thesis. I congratulate the author on his doctoral work and his thesis. As it shows, there is still much to be learned about translation and its evolution, particularly in the context of genetic code changes.

In terms of assessment, I was asked to comment on: (i) the quality of writing – this was quite fine, with a logical introduction and discussion; only minor corrections needed as detailed below; (ii) scientific contribution — substantive and more than adequate, with involvement in multiple research papers; (iii) novelty of the work — this is highly subjective, so I abstain from comment.

I note that my main scientific review of the central paper of this thesis, Kachale et al. 2023, has actually taken place in a different format from this thesis review: a separate response manuscript addressing key matters that were raised in this paper. This response is a far more careful and critical review than what I've written here, so should also be consulted. I will not reiterate these matters here. I am unsure if there's time or the ability of some of this feedback to be included in the final thesis. As far as I'm concerned, minor corrections aside, the thesis is fine as is and the decision about how to deal with this lies with the thesis author.

Main comments and suggestions:

- I would recommend including the supplement of Kachale et al. 2023, if it isn't prohibitively long, since this journal forces authors to put so much material there. There's important info in the supplement too, some of which is necessary for full evaluation of the main paper.
- The two Schultz and Yarus 1994 papers: "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)" and "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)" are both very relevant to the 4 bp tRNA AS hypothesis for UGA translation. I would thus propose adding a paragraph to the introduction or perspective explaining exactly why.

Minor comments and suggestions:

- References need to be fixed in places, with some papers cited that appear to have missing references (e.g. Pg 78 and 81), and others duplicated (e.g. Osawa et al. 1992, Pg 96 and Swart et al. 2016 Pg 100). These are just the cases I noticed.
- Pg 7: "non-overlapping" genetic code. Sometimes genes overlap, esp. in viruses. To account for this, after "means that…" add "for each gene (or mRNA)" …"each nucleotide is part of only one codon".
- Pg 7: "**a** three-nucleotide codon" (addition in **bold** font)
- Pg 7: "**Typically** each codon codes~~, however,~~ for only a single amino acid"- there are important exceptions, e.g. UGA translation as selenocysteine.
- Pg 11: "**The** reaction of translocation…"
- Pg 13: "…ochre stop signal must be represented by UAA". Citation needed.
- Pg 14: "…met with limited success so far". Citation needed.
- Pg 15: "all organisms… were found to share the same code". This is because not many organisms were examined because RNA/DNA sequencing was uncommon until Sanger sequencing was developed.
- Pg 16: "nothing inherently advantageous about the observed specific codon assignments" – Mention and discuss the counterpoint "The genetic code is one in a million" – Freeland and Hurst 1998, which advances beyond what was considered in the 1960s. Plus anything that counters Freeland and Hurst in the 20 years since this was published.
- Pg 16: (Bender et al. 2008…Sengupta and Higgs 2015). Would be better to cite original papers for the preceding statement.
- Pg 16: "reassignement" – remove 2nd "e".
- Pg 16: Mention the mechanism is explained in the next section.
- Pg 17: "two **main** theories" (other theories exist, hence "main").
- Pg 17: "Most mitochondrial genomes are small and encode ~~a lower number~~ few genes" – citation needed.
- Pg 18: "shrunk**en**"
- Pg 19: "Schultz and Yarus (1994)" – Also from earlier work published in 1993, and also the pair of "tRNA structure and ribosomal function" papers referred to in "Main comments and suggestions".
- Pg 21: "**The** same UGA-to-tryptophan reassignment…"
- Pg 21: What is the explanation for UGA->sense in Spiroplasma? It isn't evident from the context.
- Pg 22: "…UAG and UAA codons have been reassigned to cysteine and tryptophan or glutamine, respectively…" Sure about that? In eukaryotes, typically UGA=Cys/Trp, whereas UAA/UAG are Gln.
- Pg. 22: "**A** final example is represented by the ~~newly discovered~~ phylum Gricilibacteria…" (not new anymore).
- Pg 24: Legend for organism symbols needed.
- Pg 25: Phylogeny lines are too fine to be printed mostly.
- Pg 26: Cocquyt et al. 2010?
- Pg 27: "reassignement" (delete 2nd "e")
- Pg 27: "…as follow**s**:"
- Pg 27: New paragraph after "UAA and UAG stops**.**"
- Pg 27: "…reassigned via $tRNA^{Trp}CCA$ ~~or~~ **to** $tRNA^{Trp}UCA$ **editing** rely on…"
- Pg 28: "the C34 position" – needs space after "the".

- Pg 29: “flagellates”.
- Pg 30: “…where ~~does~~ the information…”
- Pg 37: “…most of them are confined to the single-celled eukaryotic lineages”. This sounds like it is somehow surprising, but most of eukaryotic molecular diversity actually lies in single-celled eukaryotes, not multicellular ones. So, one would actually expect more genetic code diversity in unicellular eukaryotes.
- Pg 37: “their complex genome**s**” (they have two nuclear genomes: germline and soma; arguably in some, but not all, ways the somatic genome is simpler than many other bloated eukaryotic genomes, like our own)
- Pg 37: “…they remain genetically intractable” – this is certainly not true: both Paramecium and Tetrahymena are molecular models with good molecular tools that are regularly used (permitting knockdowns, knockouts, expression of tagged proteins, etc.). Paramecium and Tetrahymena are excellent for classical genetics because one can easily generate completely homozygous cell lines with them. In one round of “self-fertilization” (autogamy) with Paramecium. This makes them easier to work with than fruit flies in some ways.
- Pg 41: …the ~~predicted~~ **cognate** tRNA gene was not found…” – only one cognate tRNA was not found, tRNA(UCA). The other two tRNAs corresponding to the UAA and UAG were predicted in the genome
- Pg 41: “…tRNAs cognate ~~for~~ **to** “termination” codons” – added quotes around “termination”, as they are not really termination codons anymore.
- Pg 41: “Therefore, for most non-canonical genetic codes found… tRNAs cognate… remain unknown”. This is not correct: usually tRNAs have been reported for UAA and UAG when the genomes became available; UGA is the only one typically lacking a cognate tRNA.
- Pg 41: “..of the genus Colpoda” – Heaphy et al. 2016 shows this genus uses UGA as a stop. Safest would be to just leave this out.
- Pg 41: Grimm et al. 1998 is out of date knowledge. The relevant sentences written should be replaced with Turanov et al. 2009 Science, which reports a cognate tRNA-Cys.
- Pg 41: “…four eukaryotic genera…” It’s actually more genera than this – see Seah et al. 2002.
- Pg 70: More background is needed on initiator and elongator tRNA-Met to appreciate what’s going on with T. brucei.
- Pg 70: Cite Tan et al. 2002 PNAS for the statement about the two tRNA-Met variants.
- Pg 71, first sentence: Does the T. brucei mitochondrial genome encode a tRNA-Met-i?
- Pg 72: (Fig 19) should be (Fig 23) I think. Same issue on Pg. 73.
- Pg 75: Table 5’s bold formatting seems to be missing. “combina**ti**ons”. “t”’s and “I”’s also missing subsequently, e.g. “Recogni**t**ion”.
- Pg 77: “Extended wobble ~~table~~ **schema**...”.
- Pg 78: Zoltner et al. 2018 missing from references.
- Pg 79, last sentence: NMD knockdown seems to have little effect too, e.g. Jaillon et al. 2008 Nature.
- Pg 80: “…least abundant amino acid” – citation needed.
- Pg 80: How many introns does Blastocrithidia have? Some organisms that have lost or have very divergent NMD components have very few introns, e.g. Giardia.
- Pg 81: Zurcher et al. 2002 is missing from the bibliography.
- Pg 81: The selective advantage is hypothetical until fair evidence accumulates for it, so adding “may” would indicate this. Robustness to premature termination codons was also proposed for

Condylostoma (Swart et al. 2016). Maybe the same idea was suggested in earlier publications where one or two stops were reassigned, though I cannot think of any immediately. The viral resistance hypothesis is a good selective one for alternative genetic codes in general. To date, I'm unaware of any bona fide viruses in ciliates, despite many "endosymbiotic" bacteria and algae being observed over the years. Only bacteriophages that presumably depend on the endosymbionts have been noted.

Sincerely,

Dr. Estienne Swart
(Group Leader, MPI for Biology)
16 June 2023