Karyorelict ciliates use an ambiguous genetic code with context-dependent stop/sense codons

Brandon Kwee Boon Seah, Aditi Singh, Estienne Carl Swart

Max Planck Institute for Biology, 72076 Tübingen, Germany

Correspondence:

Brandon Kwee Boon Seah, kb.seah@tuebingen.mpg.de

Estienne Carl Swart, estienne.swart@tuebingen.mpg.de

Abstract

In ambiguous stop/sense genetic codes, the stop codon(s) not only terminate translation but can also encode amino acids. Such codes have evolved at least four times in eukaryotes, twice among ciliates (*Condylostoma magnum* and *Parduczia* sp.). These have appeared to be isolated cases whose next closest relatives use conventional stop codons. However, little genomic data have been published for the Karyorelictea, the ciliate class that contains *Parduczia* sp., and previous studies may have overlooked ambiguous codes because of their apparent rarity. We therefore analyzed single-cell transcriptomes from four of the six karyorelict families to determine their genetic codes. Reassignment of canonical stops to sense codons was inferred from codon frequencies in conserved protein domains, while the actual stop codon was predicted from full-length transcripts with intact 3'-UTRs. We found that all available karyorelicts use the *Parduczia* code, where canonical stops UAA and UAG are reassigned to glutamine, and UGA encodes either tryptophan or stop. Furthermore, a small minority of transcripts may use an ambiguous stop-UAA instead of stop-UGA. Given the ubiquity of karyorelicts in marine coastal sediments, ambiguous genetic codes are not mere marginal curiosities but a defining feature of a globally distributed and abundant group of eukaryotes.

Introduction

In addition to the “standard” genetic code used by most organisms, there are numerous variant codes across the tree of life, and new ones continue to be discovered (Ivanova et al. 2014; Kollmar and Mühlhausen 2017; Shulgina and Eddy 2021). The differences lie in which amino acids are coded by which codon, as well as which codons are used to start and terminate translation (stop codons). Much of the variation is concentrated in a small number of codons, particularly the canonical stop codons UAA, UAG, and UGA, which have repeatedly been reassigned to encode amino acids. The most striking variants are ambiguous codes where one codon can have multiple meanings. This can be stochastic, such as in stop codon readthrough (Schueren and Thoms 2016), or translation of CUG as either leucine or serine by *Candida* spp. (Suzuki et al. 1997). Alternatively, they can be context-dependent, such as UGA encoding selenocysteine only in selenoproteins (Hatfield and Gladyshev 2002), meaning that the translation system is able to interpret the codon correctly as either an amino acid or a stop.

Other context-dependent stop/sense codes have been discovered where all the stop codons used by the cell are potentially also sense codons. These have evolved independently several times among the eukaryotes (Heaphy et al. 2016; Swart et al. 2016; Záhonová et al. 2016; Bachvaroff 2019): parasitic trypanosomes of the genus *Blastocrithidia* (three different species) use UAA and UAG to encode stop/glutamate (NCBI Genetic Codes ftp.ncbi.nih.gov/entrez/misc/data/gc.prt, table 31); a strain of the marine parasitic alveolate *Amoebophrya* and a marine karyorelict ciliate, *Parduczia* sp., have convergently evolved to use UGA for stop/tryptophan (table 27); and the marine heterotrich ciliate *Condylostoma magnum* uses UGA for stop/tryptophan and UAA/UAG for stop/glutamine (table 28).

The ciliates are a clade with an unusual propensity for variant genetic codes (Lozupone et al. 2001). At least eight different nuclear genetic codes are used by ciliates (Swart et al. 2016), including some of the first examples of variant codes documented in nuclear genomes (Helftenbein 1985; Horowitz and Gorovsky 1985; Preer et al. 1985; Meyer et al. 1991; Tourancheau et al. 1995). At first glance, organisms that use these ambiguous stop/sense codes appear to be isolated single species or strains embedded among relatives with conventional codes. For example, other heterotrichs related to *Condylostoma* use the standard code (e.g. *Stentor*) or the *Blepharisma* code. Additionally, a previous survey of genetic codes across the ciliate tree, including numerous uncultivated heterotrichs and karyorelicts, did not report any new examples of organisms that use ambiguous stop/sense

59 codes, nor appeared to have accounted for such a possibility in their methods (Yan et al.
60 2019). However our own preliminary studies appeared to contradict this, finding other
61 karyorelicts that use the same genetic code as *Parduczia*.

62 The karyorelicts are a class-level taxon within the ciliates, and sister group to the
63 heterotrichs. They are globally distributed and abundant in the sediment interstitial habitat of
64 marine coastal environments (Fenchel 1969). At least ~150 species have been formally
65 described but this is believed to be a severe underestimate of the true diversity (Foissner
66 1998; Ma et al. 2022), and they are also poorly represented in sequence databases.

67 We therefore sequenced additional karyorelict transcriptomes and reanalyzed published
68 data to assess whether karyorelicts other than *Parduczia* could be using ambiguous genetic
69 codes.

70 Results

71 Ten new single-cell RNA-seq libraries from karyorelicts and heterotrichs were sequenced in
72 this study, representing interstitial species from marine sediment at Roscoff, France. These
73 were analyzed alongside 33 previously published RNA-seq libraries (Table S1). After filtering
74 for quality and sufficient data, 25 transcriptome assemblies were used to evaluate stop
75 codon reassignment (15 previously published), vs. 26 assemblies (16 previously published)
76 for inferring the actual stop codon(s) (Supplementary Information).

77 *Reassignment of all three canonical stop codons to sense codons in karyorelicts*

78 Codon frequencies in protein-coding sequences were calculated from sequence regions that
79 aligned to conserved Pfam domains, in transcripts with poly-A tails. Transcriptomes and
80 genomic coding sequences (CDSs) from ciliates with known genetic codes were used as a
81 comparison to estimate the false positive rate of stop codons being found in these
82 alignments, e.g. because of misalignments, misassembly, or pseudogenes.

83 Among karyorelicts, all three canonical stop codons (UAA, UAG, UGA) were observed in
84 conserved protein domains, with frequencies between 0.08-2.9%, which fell within the range
85 observed for coding codons in organisms with known genetic codes (0.03-6.8%, excluding
86 the outlier CGG in *Tetrahymena thermophila* with only 0.003%). This range was also similar
87 to frequencies of the ambiguous stops in *Parduczia* and the heterotrich *Condylostoma*
88 (Figure 1). UGA was generally less frequent than UAA/UAG in all karyorelicts, but the
89 frequencies varied between taxa, reflecting their individual codon usage biases or which

90 genes are assembled in the transcriptome because of sequencing depth. UGA was the
91 least-frequent codon in most Trachelocercidae and Geleiidae, but was more frequent in
92 Loxodidae and Kentrophoridae than some other codons, especially C/G-rich ones like CGG
93 (Figure 1). Nonetheless, these were all still one to two orders of magnitude higher than the
94 observed frequencies of in-frame actual stops from other ciliate species in the reference set.

95 In-frame UGAs were found in 10.5 to 76.9% of transcripts with putative coding regions
96 predicted by full-length Blastx hits per karyorelict library (Figure 4D). This verified that in-
97 frame UGAs were not concentrated in a small fraction of potentially spurious sequences but
98 in fact found in many genes. Conserved 'marker' genes that were generally expected to be
99 present in ciliate genomes (BUSCO orthologs, Alveolata marker set) also contained in-frame
100 UGAs. The karyorelict transcriptome assemblies were relatively incomplete, with 1.8% to
101 20.5% (median 12.0%) estimated completeness based on the BUSCO markers, and a total
102 of 91 of 171 BUSCO orthologs were found in these assemblies (Figure 2A). Nonetheless, 46
103 BUSCO orthologs from 14 karyorelict assemblies were found with in-frame UGAs in
104 conserved alignment positions (e.g. Figure 2B, 2C), verifying that they are not limited to
105 poorly characterized or hypothetical proteins.

106 In comparison, the heterotrich *Anigsteinia*, for which two new sequence libraries were also
107 produced and which was found in the same habitats as karyorelicts, had in-frame
108 frequencies of ≤0.011% for all three canonical stop codons, which were comparable to
109 frequencies of the known stop codons in *Blepharisma* and *Stentor* (max. 0.09%). Hence
110 *Anigsteinia* probably does not have ambiguous sense/stop codons.

111 All karyorelicts had the same inferred amino acid reassignments for the three canonical
112 stops: glutamine (Q) for UAA and UAG, and tryptophan (W) for UGA (Figure 3), matching
113 previous predictions for *Parduczia* sp. and *Condylostoma magnum* (Heaphy et al. 2016;
114 Swart et al. 2016).

115 *Stop codons in karyorelicts and heterotrichs*

116 Frequency of a codon in coding regions can be used to infer if it is a sense codon but not
117 whether it can terminate translation, especially for ambiguous codes where codons that can
118 terminate translation also frequently appear in coding sequences. Therefore we used full
119 length transcripts with both a high quality Blastx alignment to a reference protein and a poly-
120 A tail to predict the likely stop codon(s) used in each sample. To avoid double counting, only
121 one isoform was used per gene. We assumed that the true stop codon(s) were one or more
122 of the three canonical stops UGA, UAA, UAG, and that if a contig has a high quality Blastx

123 hit to a reference protein sequence, the true stop should lie somewhere between the last
124 codon at the 3' end of the hit region and the beginning of the poly-A. We reasoned that if the
125 true stop codon set was used for annotation, (i) the number of transcripts without a putative
126 true stop should be minimized; (ii) the variance of the 3'-UTR length should also be
127 minimized because ciliate 3'-UTRs are known to be short (mostly <100 bp); and (iii) if there
128 was more than one stop codon, the length distributions of the putative 3'-UTRs for each stop
129 codon should be centered on the same value.

130 With these criteria, the candidate stop codons for karyorelicts could be narrowed to two
131 possibilities: UGA alone or UGA + UAA. If only UGA was permitted as a stop codon, 84-98%
132 of transcripts per library had a putative true stop, but if both UGA and UAA were permitted
133 as stop codons, the proportion was over 98% (Figure 4A). This was at the expense of
134 somewhat more variance in the 3'-UTR length distribution, although we found that this metric
135 was of limited usefulness because UGA was always the majority in all stop codon
136 combinations where it was present (Figures 4B, 4C). Both karyorelicts and heterotrichs in
137 this study had short and narrowly distributed 3'-UTR lengths (median 28 nt, interquartile
138 range 18 nt) (Figure 4C). The heterotrichs were shortest overall, with median lengths per
139 taxon between 21 nt (*Condylostoma*) and 26 nt (*Stentor*), followed by the karyorelict families
140 Trachelocercidae (33 nt), Geleiidae (31 nt), Kentrophoridae (37 nt), and Loxodidae (43 nt).

141 In previous analyses of the ambiguous stop codons in *Condylostoma* and *Parduczia*, a
142 distinct depletion of in-frame coding "stop" codons immediately upstream of the actual
143 terminal stop was observed (Swart et al. 2016). We could reproduce this depletion of all
144 three canonical stops in *Condylostoma* and of UGA in *Parduczia*, about 10 to 20 codon
145 positions before the putative terminal stop, in our reanalysis of the same data (Figure 5A).
146 For the karyorelicts, if only UGA was permitted as a stop codon, we observed depletion of
147 coding-UGA but also of coding-UAAs before the terminal stop-UGA (Figure 5B). If UGA +
148 UAA were permitted as stops, the depletion of coding-UGA before terminal stops was still
149 observed, while the depletion of coding-UAA was even more pronounced (Figure 5C).
150 Unfortunately, there were only a limited number of full-length karyorelict transcripts with
151 putative stop-UAAs (max. 47 contigs per library). We, therefore, concluded that UGA is the
152 predominant stop codon in karyorelicts, but UAA may also function as a stop codon for about
153 1-10% of transcripts.

154 UAA and UAG were predicted as stop codons of *Anigsteinia* (Spirostomidae), consistent with
155 their near-absence from coding regions in this genus (see above, Figure 1). UGA was not
156 only near-absent from coding regions, but also rarely encountered as a putative stop codon,

157 although it was not uncommon in 3'-UTRs. Similar rarity of UGAs as putative stops was also
158 observed in *Stentor* and other heterotrichs that are said to use the standard code. Either (i)
159 these heterotrichs use the standard genetic code with all three canonical stop codons but a
160 strong bias against using UGA for stop, or (ii) UGA is an unassigned codon in these
161 organisms.

162 Discussion

163 We have found that the codon UGA is used as both a stop codon and to code for tryptophan
164 by karyorelictean ciliates. The taxa sampled represent four of the six families of karyorelicts:
165 Loxodidae, Trachelocercidae, Geleiidae, and Kentrophoriidae. When this distribution is
166 mapped to an up-to-date phylogeny (Ma et al. 2022), we can infer that this ambiguous code,
167 formerly reported only for *Parduczia* sp. (Geleiidae) among ciliates, was actually acquired at
168 the root of the karyorelict clade (Figure 6).

169 Available data for *Cryptopharynx* (Karyorelictea: Cryptopharyngidae) were not conclusive.
170 The canonical stop codons had frequencies between 0.02 and 0.07%, lower than for other
171 karyorelicts, but higher than true stop codons, but this family was represented by a single
172 library that had high contamination from other eukaryotes (Figure S1) and there were too few
173 high-confidence, full length transcripts for a reliable conclusion on its genetic code. No
174 sequence data beyond rRNA genes were publicly available for the remaining family,
175 Wilbertomorphidae, whose phylogenetic position in relation to the other karyorelicts is
176 unclear because of long branch lengths (Xu et al. 2013).

177 Ambiguous stop/sense codes are hence not just isolated phenomena, but are used by a
178 major taxon that is diverse, globally distributed, and abundant in its respective habitats. In
179 contrast, the heterotrichs, which constitute the sister group to Karyorelictea and are hence of
180 the same evolutionary age, use at least three different genetic codes, including one with
181 ambiguous stops (Figure 6). If organisms with ambiguous codes were isolated single
182 species whose nearest relatives have conventional stops, as appears to be the case for
183 *Blastocrithidia* spp. and *Amoebophrya* sp., we might conclude that these are uncommon
184 occurrences that do not persist over longer evolutionary time scales. However, the
185 karyorelict crown group diversified during the Proterozoic (posterior mean 455 Mya) and the
186 stem split from the Heterotrichea even earlier, in the Neo-Proterozoic (Fernandes and
187 Schrago 2019).

188 This study has benefited from several technical improvements. A highly complete,

189 contiguous genome assembly with gene predictions is now available for the heterotrich
190 *Blepharisma stoltei* (M. Singh et al., BioRxiv doi:10.1101/2021.12.14.471607). Because
191 *Blepharisma* is more closely related to the karyorelicts than other ciliate model species,
192 which are mostly oligohymenophorans and spirotrichs, it improved the reference-based
193 annotation of the assembled transcriptomes. Single-cell RNA-seq libraries in this study were
194 also sequenced to a greater depth, with a lower fraction of contamination from rRNA, and
195 hence yielded more full length mRNA transcripts for analysis.

196 One proposed mechanism for how the cell correctly recognizes whether an ambiguous
197 codon is coding or terminal is based on the proximity of translation stops to the poly-A tail of
198 transcripts. In this model, tRNAs typically bind more efficiently to in-frame coding "stops"
199 than eukaryotic translation release factor 1 (eRF1), hence allowing these codons to be
200 translated. At the true termination stop codon, however, the binding of eRF1 can be
201 stabilized by interactions with poly-A interacting proteins like PABP bound to the nearby
202 poly-A tail, allowing it to outcompete tRNAs and hydrolyze the peptidyl-tRNA bond (Swart et
203 al. 2016; Alkalaeva and Mikhailova 2017). Consistent with this model, we found that
204 karyorelict 3'-UTRs are also relatively short, and that in-frame UGAs are depleted
205 immediately before the putative true stop codon. Nonetheless, karyorelict 3'-UTRs are
206 actually about 10 nt longer on average than those of heterotrichs.

207 Our results also raised the possibility that UAA is also used as an ambiguous stop codon for
208 ~1-10% of karyorelict transcripts, in addition to the main stop codon UGA. eRF1 may retain
209 a weak affinity for UAA, and recognize UAA for terminating translation albeit with lower
210 efficiency. In *Blepharisma japonicum*, where UAA and UAG are non-ambiguous stops and
211 UGA encodes tryptophan (albeit at low frequency, 0.13%), heterologously expressed eRF1
212 could still recognize all three codons in an *in vitro* assay, although efficiency of peptidyl-tRNA
213 hydrolysis was lower with UGA than for UAA and UAG (Eliseev et al. 2011). In species with
214 non-ambiguous stop codon reassignment, the effect of such "weak" ambiguity on the total
215 pool of translated protein may be negligible, but it shows that there is a latent potential that
216 could account for the repeated evolution of stop codon reassignments in ciliates.
217 Furthermore, UAAs were even more abundant than UGAs in ciliate 3'-UTRs, which can be
218 attributed to the low GC% of 3'-UTRs compared to coding sequences; other A/U-only
219 codons were also enriched in 3'-UTRs. Therefore, UAAs in the 3'-UTRs of karyorelicts may
220 be a 'backstop' mechanism that prevents occasional stop-codon readthrough, as proposed
221 for tandem stop codons (TSCs) in other species with reassigned stop codons (Fleming and
222 Cavalcanti 2019). In the minority of transcripts where in-frame stop-UGA is absent, the

backstop may be adequate to terminate translation before the poly-A tail and produce a functional protein most of the time.

What selective pressures might favor the evolution and maintenance of an ambiguous genetic code? One possibility is that they confer mutational robustness by eliminating substitutions that cause premature stop codons. They do not appear to be linked to a specific habitat: *Blastocrithidia* spp. and *Amoebophrya* sp. are both parasites of eukaryotic hosts, but of insects and free-living dinoflagellates respectively; whereas the karyorelict ciliates and *Condylostoma* are both found in marine interstitial environments, but live alongside other ciliates that have conventional codes, such as *Anigsteinia*. Having short 3'-UTRs may predispose ciliates to adopt ambiguous codes by facilitating interactions between eRF1 and PABPs that could enable stop recognition, but it is not the only deciding factor because the 3'-UTRs of ciliates with conventional stop codons are also comparably short, particularly among the heterotrichs.

Any adaptationist hypothesis for alternative and ambiguous codes will have to contend with the existence of related organisms with conventional codes that have similar lifestyles. Furthermore, once a stop codon has been reassigned to sense, it becomes increasingly difficult to undo without the deleterious effects of premature translation termination, and may function like a ratchet. Like the origins of the genetic code itself (Crick 1968), we may have to be content with the null hypothesis that they are “frozen accidents” that reached fixation stochastically, and which are maintained because they do not pose a significant selective disadvantage.

Materials and Methods

Sample collection

Surface sediment was sampled in September 2021 from two sites in the bay at Roscoff, France when exposed at low tide. Site A: shallow swimming enclosure, 48.72451 N, 3.992294 W; Site B: adjacent to green algae tufts near freshwater outflow, 48.716169 N, 3.995626 W. Upper 1-2 cm of sediment was skimmed into glass beakers, and stored under local seawater until use. Interstitial ciliates were collected by decantation: a spoonful of sediment was stirred in seawater in a beaker. Sediment particles were briefly allowed to settle out, and the overlying suspended organic material was decanted into Petri dishes. Ciliate cells were identified by morphology under a dissection microscope and picked by pipetting with sterile, filtered pipette tips. Selected cells were imaged with incident light under

255 a stereo microscope (Olympus SZX10, Lumenera Infinity 3 camera).

256 NEBNext cell lysis buffer (NEB, E5530S) was premixed and filled into PCR tubes; per tube:
257 0.8 µL 10x cell lysis buffer, 0.4 µL murine RNAse inhibitor, 5.3 µL nuclease-free water.
258 Picked ciliate cells were transferred twice through filtered local seawater (0.22 µm, Millipore
259 SLGP033RS) to wash, then transferred with 1.5 µL carryover volume to 6.5 µL of cell lysis
260 buffer (final volume 8 µL), and snap frozen in liquid nitrogen. Samples were stored at -80 °C
261 before use.

262 *Single-cell RNAseq sequencing*

263 Samples collected in cell lysis buffer (Table S1) were used for RNAseq library preparation
264 with the NEBNext Single Cell / Low Input RNA Library Prep Kit for Illumina (NEB, E6420S),
265 following the manufacturer's protocol for single cells, with the following parameters adjusted:
266 17 cycles for cDNA amplification PCR, cDNA input for library enrichment normalized to 3 ng
267 (or all available cDNA used for libraries where total cDNA was <3 ng), 8 cycles for library
268 enrichment PCR. Libraries were dual-indexed (NEBNext Dual Index Primers Set 1, NEB
269 E7600S), and sequenced on an Illumina NextSeq 2000 instrument with P3 300 cycle
270 reagents, with target yield of 10 Gbp per library.

271 *RNA-seq library quality control and transcriptome assembly*

272 Previously published karyorelict transcriptome data (Keeling et al. 2014; Kolisko et al. 2014;
273 Seah et al. 2019; Yan et al. 2019) were downloaded from the European Nucleotide Archive
274 (ENA) (Table S1). Contamination from non-target organisms was evaluated by mapping
275 reads to an rRNA reference database and summarizing the hits by taxonomy. Although
276 RNAseq library construction enriches mRNAs using poly-A tail selection, there is typically
277 still sufficient rRNA present in the final library to evaluate the taxonomic composition of the
278 sample. All RNAseq read libraries (newly sequenced and previously published) were
279 processed with the same pipeline: The taxonomic composition of each library was evaluated
280 by mapping 1 M read pairs per library against the SILVA SSU Ref NR 132 database (Quast
281 et al. 2013), using phyloFlash v3.3b1 (Gruber-Vodicka et al. 2020). Newly sequenced
282 libraries were assigned to a genus or family using the mapping-based taxonomic summary,
283 or full-length 18S rRNA gene if it was successfully assembled.

284 Reads were trimmed with the program bbduk.sh from BBmap v38.22 (http://sourceforge.net/
285 projects/bbmap/) to remove known adapters (right end) and low-quality bases (both ends),
286 with minimum Phred quality 24 and minimum read length 25 bp. Trimmed reads were then

287 assembled with Trinity v2.12.0 (Grabherr et al. 2011) using default parameters. Assembled
288 contigs were aligned against the *Blepharisma stoltei* ATCC 30299 proteome (M. Singh et al.,
289 BioRxiv doi:10.1101/2021.12.14.471607) with NCBI Blastx v2.12.0 (Camacho et al. 2009)
290 using the standard genetic code and E-value cutoff 10^{-20}, parallelized with GNU Parallel
291 (Tange 2018).

292 Read pre-processing, quality control, and assembly were managed with a Snakemake
293 v6.8.1 (Mölder et al. 2021) workflow (https://github.com/Swart-lab/karyocode-workflow).
294 Scripts for data processing described below were written in Python v3.7.3 using Biopython
295 v1.74 (Cock et al. 2009), pandas v0.25.0 (McKinney 2010), seaborn v0.11.0 (Waskom 2021)
296 and Matplotlib v3.1.1 (Hunter 2007) libraries unless otherwise stated.

297 *Prediction of stop codon reassignment to sense*

298 Only contigs with poly-A tails ≥7 bp were used for genetic code prediction, to exclude
299 potential bacterial contaminants, especially because several species (*Kentrophoros* spp.,
300 *Parduczia* sp., Figure S1) are known to have abundant bacterial symbionts. Presence and
301 lengths of poly-A tails in assembled transcripts were evaluated with a Python regular
302 expression. Library preparation was not strand-specific, hence contigs starting with poly-T
303 were reverse-complemented, and contigs with both a poly-A tail and a poly-T head
304 (presumably fused contig) were excluded.

305 Codon frequencies and their corresponding amino acids were predicted with an updated
306 version of PORC (v2.1, https://github.com/Swart-lab/PORC, managed with a Snakemake
307 workflow, https://github.com/Swart-lab/karyocode-analysis-porc); the method has been
308 previously described (Dutilh et al. 2011; Swart et al. 2016). Briefly: a six-frame translation
309 was produced for each contig in the transcriptome assembly, and searched against
310 conserved domains in the Pfam-A database v32 (Mistry et al. 2021) with hmmscan from
311 HMMer v3.3.2 (http://hmmer.org/). Overall codon frequencies were counted from alignments
312 with E-value ≤ 10^{-20}. To ensure that there was sufficient data underlying the codon and
313 amino acid frequencies, only those libraries with at least 100 observations for each of the
314 coding codons in the standard genetic code were used for comparison of codon frequencies
315 and for prediction of amino acid assignments.

316 Frequencies of amino acids aligning to a given codon were counted from columns where the
317 HMM model consensus was ≥50% identity in the alignment used to build the model (upper-
318 case positions in the HMM consensus). Sequence logos of amino acid frequencies per
319 codon for each library were drawn with Weblogo v3.7.5 (Crooks et al. 2004).

320 In addition to the transcriptomes, genomic CDSs of selected model species with different
321 genetic codes (Aury et al. 2006; Stover et al. 2006; Slabodnick et al. 2017; Arnaiz et al.
322 2020; Sheng et al. 2020) (M. Singh et al., BioRxiv doi:10.1101/2021.12.14.471607) were
323 also analyzed with PORC to obtain a reference baseline of coding-codon frequencies (Table
324 S2). These model species have non-ambiguous codes so they were not expected to have
325 stop codons in the CDSs, except for the terminal stop.

326 *Prediction of coding frame in full-length transcripts*

327 "Full-length" transcripts (with poly-A tail, intact 3'-UTR, and complete coding sequence) were
328 desirable to predict the stop codon, characterize 3'-UTR metrics, and verify genetic code
329 predictions. Contigs were therefore filtered with the following criteria: (i) poly-A tail ≥7 bp,
330 criterion following (Swart et al. 2016), (ii) contig contains a Blastx hit vs. *B. stoltei* protein
331 sequence with E-value $\leq 10^{-20}$ and where the alignment covers ≥80% of the reference *B.*
332 *stoltei* sequence, (iii) both poly-A tail and Blastx hit agree on the contig orientation. For
333 contigs with multiple isoforms assembled by Trinity, the isoform with the longest Blastx hit
334 was chosen; in case of a Blastx hit length tie, then the longer isoform was chosen. Only
335 libraries with >100 assembled "full-length" transcripts were used for downstream analyses
336 (Figure S2).

337 *Metrics for evaluating potential stop codon combinations*

338 For each of the 7 possible combinations of the 3 canonical stop codons (UGA, UAA, UAG),
339 we treated the first in-frame stop downstream of the Blastx hit in each full-length transcript
340 (including the last codon of the hit) as the putative stop codon, and recorded the number of
341 full-length transcripts with a putative stop, the length of the 3'-UTR (distance from stop to
342 beginning of the poly-A tail), as well as the codon frequencies for each position from 150
343 codons upstream of the putative stop to the last in-frame three-nucleotide triplet before the
344 poly-A tail.

345 *Delimitation of putative coding sequences using Blastx hits*

346 The start codon was more difficult to evaluate because the 5' end of the transcript may not
347 have been fully assembled, and there was no straightforward way to recognize its
348 boundaries, unlike the 3'-poly-A tail. We used the following heuristic criteria to define the
349 start of the CDS: first in-frame ATG upstream of the Blastx hit (including first codon of the
350 hit), or first in-frame stop codon encountered upstream (to avoid potential problems with
351 ORFs containing in-frame stops), whichever comes first. Otherwise, the transcript was

352 assumed to be incomplete at the 5'-end and simply truncated with the required 1 or 2 bp
353 offset to keep the CDS in frame.

354 *Verification of in-frame UGAs in conserved marker genes*
355 Full-length CDSs (see above) were translated with the karyorelict code (NCBI table 27).
356 Conserved marker genes were identified with BUSCO v5.2.2 (protein mode,
357 alveolata_odb10 marker set) (Waterhouse et al. 2018), managed with a Snakemake
358 workflow (https://github.com/Swart-lab/karyocode-analysis-busco). Markers for additional
359 ciliate species where relatively complete genome assemblies and gene predictions were
360 available were also identified (Table S3) (Coyne et al. 2011; Swart et al. 2013; Aeschlimann
361 et al. 2014; McGrath et al. 2014; Xiong et al. 2015; Hamilton et al. 2016; Arnaiz et al. 2017;
362 Slabodnick et al. 2017; Chen et al. 2019; Sheng et al. 2020). For each BUSCO marker, the
363 ciliate homologs were aligned with Muscle v3.8.1551 (Edgar 2004). Alignment columns
364 corresponding to in-frame putatively coding UGAs of karyorelict sequences were identified.
365 These positions were considered to be conserved if ≥50% of residues were W or another
366 aromatic amino acid (Y, F, or H).

367 Figure legends

368 **Figure 1.** Codon frequencies of canonical stop codons (UGA: blue, UAA: orange, UAG:
369 green) and other codons (gray) in conserved protein domains found by hmmscan search in
370 six-frame translations of transcriptome assemblies (Table S1) or genomic CDSs (Table S2)
371 vs. Pfam. (**left**) Assignments of canonical stops for organisms with known genetic codes,
372 following Swart et al., 2016. Names of libraries from this study are highlighted in bold.

373 **Figure 2.** In-frame coding UGAs in conserved marker genes. (**A**) Completeness estimates of
374 heterotrich and karyorelict transcriptomes (library names in green and blue respectively),
375 compared with genomic reference sequences from other ciliates (Table S3); BUSCO
376 Alveolata marker set. (**B, C**) Two examples of alignments (excerpts) for conserved
377 orthologous protein-coding genes (orthologs 20320at33630 and 23778at33630), which
378 contain in-frame UGAs translated as W in karyorelict sequences, flanked by conserved
379 alignment blocks.

380 **Figure 3.** Weblogos representing the likely amino acid assignment of each codon in selected
381 libraries (library with most coverage per taxon of interest). Heights of each letter represent
382 the relative frequencies (all scaled to 100%) of each amino acid in conserved residues
383 aligning to that codon. The observed codon frequency (in %) is indicated below. Codons with

384 frequencies <0.02% are highlighted in red, representing either non-ambiguous stops or
385 unassigned codons. Assignment of cysteine (C) for UGA in *Anigsteinia* is based on only 16
386 alignments, of which 14 are to a likely selenoprotein (Pfam domain GSHPx).

387 **Figure 4.** Effect of different stop codon combinations on assembly metrics. Predicted stop
388 codon usage for each taxon from this study or previous publications highlighted in gray. (**A**)
389 Strip plots for the fraction of full length contigs per transcriptome that have a putative stop
390 codon from that specific combination (rows), i.e. in-frame, downstream of full-length Blastx
391 hit vs. reference, and upstream of poly-A tail. Each point corresponds to one transcriptome
392 assembly, grouped by taxonomic family (columns). (**B**) Scatterplots for standard deviation of
393 3'-UTR lengths. (**C**) Histograms for 3'-UTR lengths, colored by putative stop codon (UGA:
394 blue, UAA: orange, UAG: green), one representative library per family. (**D**) Fraction of full-
395 length transcripts that have at least one canonical stop codon in the putative coding region,
396 grouped by family (further split to genus for Loxodidae and Geleiidae).

397 **Figure 5.** Depletion of in-frame coding "stop" codons in the coding sequence (negative
398 coordinates) immediately before the putative true stop codon (position 0) and their
399 enrichment in the 3'-UTR (positive coordinates). Representative library with highest number
400 of assembled full length contigs chosen per taxon. (**A**) Codon counts for UGA (blue), UAA
401 (orange), and UAG (green) before and after putative true stop in *Condylostoma magnum*
402 (uses all three as ambiguous stops), and three heterotrichs with unambiguous stops. (**B**)
403 Codon counts for karyorelicts if only UGA is permitted as a stop codon. (**C**) Codon counts for
404 karyorelicts if both UGA and UAA are permitted as stop codons.

405 **Figure 6.** Genetic code diversity among karyorelict and heterotrich ciliates. (**Left**)
406 Diagrammatic karyorelict + heterotrich tree with predicted stop codon reassignments
407 mapped to each family. Subtree topologies are from Ma et al. (2022) and Fernandes et al.
408 (2016) respectively. Branch lengths are not representative of evolutionary distances. (**Right**)
409 Photomicrographs of ciliates (incident light) collected in this study from Roscoff, France;
410 height of each panel 50 µm.

411 Data availability

412 RNA-seq libraries sequenced for this study have been deposited at the European Nucleotide
413 Archive under accession PRJEB50648.

Acknowledgements

This work was supported by the Max Planck Society. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 730984, Assemble Plus project. We thank S. Booker, C. Cabresin, E. Bourrigaud, and R. Garnier for facilitating our research visit to the Station Biologique Roscoff; H. Budde for sequencing; H. Gruber-Vodicka for loan of equipment; A. Peters for helpful suggestions and cake; A. Lemper for administrative assistance; Y. Shulgina for thoughtful feedback on the manuscript, and members of the Swart Lab for helpful feedback and support.

423 ## References

424 Aeschlimann SH, Jönsson F, Postberg J, Stover NA, Petera RL, Lipps H-J, Nowacki M,
425 Swart EC. 2014. The draft assembly of the radically organized *Stylonychia lemnae*
426 macronuclear genome. Genome Biol. Evol. 6:1707–1723.

427 Alkalaeva E, Mikhailova T. 2017. Reassigning stop codons via translation termination: How a
428 few eukaryotes broke the dogma. Bioessays 39.

429 Arnaiz O, Meyer E, Sperling L. 2020. ParameciumDB 2019: integrating genomic data across
430 the genus for functional and evolutionary biology. Nucleic Acids Res. 48:D599–D605.

431 Arnaiz O, Van Dijk E, Bétermier M, Lhuillier-Akakpo M, de Vanssay A, Duharcourt S, Sallet
432 E, Gouzy J, Sperling L. 2017. Improved methods and resources for *Paramecium*
433 genomics: transcription units, gene annotation and gene expression. BMC Genomics
434 18:483.

435 Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard
436 V, Aiach N, et al. 2006. Global trends of whole-genome duplications revealed by the
437 ciliate *Paramecium tetraurelia*. Nature 444:171–178.

438 Bachvaroff TR. 2019. A precedented nuclear genetic code with all three termination codons
439 reassigned as sense codons in the syndinean *Amoebophrya* sp. ex *Karlodinium*
440 *veneficum*. PLoS ONE 14:e0212912.

441 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
442 BLAST+: architecture and applications. BMC Bioinformatics 10:421.

443 Chen X, Jiang Y, Gao F, Zheng W, Krock TJ, Stover NA, Lu C, Katz LA, Song W. 2019.
444 Genome analyses of the new model protist *Euplotes vannus* focusing on genome
445 rearrangement and resistance to environmental stressors. Mol. Ecol. Resour. 19:1292–
446 1308.

447 Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T,
448 Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for
449 computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423.

450 Coyne RS, Hannick L, Shanmugam D, Hostetler JB, Brami D, Joardar VS, Johnson J,
451 Radune D, Singh I, Badger JH, et al. 2011. Comparative genomics of the pathogenic
452 ciliate *Ichthyophthirius multifiliis*, its free-living relatives and a host species provide
453 insights into adoption of a parasitic lifestyle and prospects for disease control. Genome
454 Biol. 12:R100.

455 Crick FH. 1968. The origin of the genetic code. J. Mol. Biol. 38:367–379.

456 Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo
457 generator. Genome Res. 14:1188–1190.

458 Dutilh BE, Jurgelenaite R, Szklarczyk R, van Hijum SAFT, Harhangi HR, Schmid M, de Wild
459 B, Françoijs K-J, Stunnenberg HG, Strous M, et al. 2011. FACIL: Fast and Accurate
460 Genetic Code Inference and Logo. Bioinformatics 27:1929–1933.

461 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
462 throughput. Nucleic Acids Res. 32:1792–1797.

463 Eliseev BD, Alkalaeva EZ, Kryuchkova PN, Lekomtsev SA, Wang W, Liang A-H, Frolova
464 LYu. 2011. Translation termination factor eRF1 of the ciliate *Blepharisma japonicum*
465 recognizes all three stop codons. Mol Biol (NY) 45:614–618.

466 Fenchel T. 1969. The ecology of marine microbenthos IV. Structure and function of the
467 benthic ecosystem, its chemical and physical factors and the microfauna commuities
468 with special reference to the ciliated protozoa. Ophelia 6:1–182.

469 Fernandes NM, Schrago CG. 2019. A multigene timescale and diversification dynamics of
470 Ciliophora evolution. Mol. Phylogenet. Evol. 139:106521.

471 Fleming I, Cavalcanti ARO. 2019. Selection for tandem stop codons in ciliate species with
472 reassigned stop codons. PLoS ONE 14:e0225804.

473 Foissner W. 1998. The karyorelictids (Protozoa: Ciliophora), a unique and enigmatic
474 assemblage of marine, interstitial ciliates: a review emphasizing ciliary patterns and
475 evolution. Evolutionary relationships among Protozoa:305–325.

476 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
477 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-
478 Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

479 Gruber-Vodicka HR, Seah BKB, Pruesse E. 2020. phyloFlash: Rapid Small-Subunit rRNA
480 Profiling and Targeted Assembly from Metagenomes. mSystems 5.

481 Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M,
482 Krishnakumar V, Badger JH, Caler EV, et al. 2016. Structure of the germline genome of
483 *Tetrahymena thermophila* and relationship to the massively rearranged somatic
484 genome. eLife 5.

485 Hatfield DL, Gladyshev VN. 2002. How selenium has altered our understanding of the
486 genetic code. Mol. Cell. Biol. 22:3565–3576.

487 Heaphy SM, Mariotti M, Gladyshev VN, Atkins JF, Baranov PV. 2016. Novel Ciliate Genetic
488 Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons
489 in *Condylostoma magnum*. Mol. Biol. Evol. 33:2885–2889.

490 Helftenbein E. 1985. Nucleotide sequence of a macronuclear DNA molecule coding for
491 alpha-tubulin from the ciliate *Stylonychia lemnae*. Special codon usage: TAA is not a
492 translation termination codon. Nucleic Acids Res. 13:415–433.

493 Horowitz S, Gorovsky MA. 1985. An unusual genetic code in nuclear genes of *Tetrahymena*.
494 Proc Natl Acad Sci USA 82:2452–2455.

495 Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 9:90–95.

496 Ivanova NN, Schwientek P, Tripp HJ, Rinke C, Pati A, Huntemann M, Visel A, Woyke T,
497 Kyrpides NC, Rubin EM. 2014. Stop codon reassignments in the wild. Science 344:909–
498 913.

499 Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV,
500 Archibald JM, Bharti AK, Bell CJ, et al. 2014. The Marine Microbial Eukaryote
501 Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of
502 eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol.
503 12:e1001889.

504 Kolisko M, Boscaro V, Burki F, Lynn DH, Keeling PJ. 2014. Single-cell transcriptomics for
505 microbial eukaryotes. Curr. Biol. 24:R1081-2.

506 Kollmar M, Mühlhausen S. 2017. Nuclear codon reassignments in the genomics era and
507 mechanisms behind their evolution. Bioessays 39.

508 Lozupone CA, Knight RD, Landweber LF. 2001. The molecular basis of nuclear genetic code
509 change in ciliates. Curr. Biol. 11:65–74.

510 Ma M, Li Y, Maurer-Alcalá XX, Wang Y, Yan Y. 2022. Deciphering phylogenetic relationships
511 in class Karyorelictea (Protista, Ciliophora) based on updated multi-gene information
512 with establishment of a new order Wilbertomorphida n. ord. Mol. Phylogenet.
513 Evol.:107406.

514 McGrath CL, Gout J-F, Doak TG, Yanagi A, Lynch M. 2014. Insights into three whole-
515 genome duplications gleaned from the *Paramecium caudatum* genome sequence.
516 Genetics 197:1417–1428.

517 McKinney W. 2010. Data structures for statistical computing in python. In: Proceedings of
518 the 9th Python in Science Conference. Proceedings of the python in science

519 conference. SciPy. p. 56–61.

520 Meyer F, Schmidt HJ, Plümper E, Hasilik A, Mersmann G, Meyer HE, Engström A,
521 Heckmann K. 1991. UGA is translated as cysteine in pheromone 3 of *Euplotes*
522 *octocarinatus*. Proc Natl Acad Sci USA 88:3758–3761.

523 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto
524 SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database
525 in 2021. Nucleic Acids Res. 49:D412–D419.

526 Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S,
527 Twardziok SO, Kanitz A, et al. 2021. Sustainable data analysis with Snakemake.
528 F1000Res. 10:33.

529 Preer JR, Preer LB, Rudman BM, Barnett AJ. 1985. Deviation from the universal code
530 shown by the gene for surface protein 51A in *Paramecium*. Nature 314:188–190.

531 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.
532 2013. The SILVA ribosomal RNA gene database project: improved data processing and
533 web-based tools. Nucleic Acids Res. 41:D590-6.

534 Schueren F, Thoms S. 2016. Functional translational readthrough: A systems biology
535 perspective. PLoS Genet. 12:e1006196.

536 Seah BKB, Antony CP, Huettel B, Zarzycki J, Schada von Borzyskowski L, Erb TJ, Kouris A,
537 Kleiner M, Liebeke M, Dubilier N, et al. 2019. Sulfur-Oxidizing Symbionts without
538 Canonical Genes for Autotrophic CO2 Fixation. MBio 10:e01112-19.

539 Sheng Y, Duan L, Cheng T, Qiao Y, Stover NA, Gao S. 2020. The completed macronuclear
540 genome of a model ciliate *Tetrahymena thermophila* and its application in genome
541 scrambling and copy number analyses. Sci. China Life Sci. 63:1534–1542.

542 Shulgina Y, Eddy SR. 2021. A computational screen for alternative genetic codes in over
543 250,000 genomes. eLife 10.

544 Slabodnick MM, Ruby JG, Reiff SB, Swart EC, Gosai S, Prabakaran S, Witkowska E, Larue
545 GE, Fisher S, Freeman RM, et al. 2017. The Macronuclear Genome of *Stentor*
546 *coeruleus* Reveals Tiny Introns in a Giant Cell. Curr. Biol. 27:569–575.

547 Stover NA, Krieger CJ, Binkley G, Dong Q, Fisk DG, Nash R, Sethuraman A, Weng S,
548 Cherry JM. 2006. Tetrahymena Genome Database (TGD): a new genomic resource for
549 Tetrahymena thermophila research. Nucleic Acids Res. 34:D500-3.

550 Suzuki T, Ueda T, Watanabe K. 1997. The “polysemous” codon--a codon with multiple

551 amino acid assignment caused by dual specificity of tRNA identity. EMBO J. 16:1122–
552 1134.

553 Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD,
554 Nowacki M, Schotanus K, et al. 2013. The *Oxytricha trifallax* macronuclear genome: a
555 complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol. 11:e1001473.

556 Swart EC, Serra V, Petroni G, Nowacki M. 2016. Genetic Codes with No Dedicated Stop
557 Codon: Context-Dependent Translation Termination. Cell 166:691–702.

558 Tange O. 2018. Gnu Parallel 2018. Zenodo.

559 Tourancheau AB, Tsao N, Klobutcher LA, Pearlman RE, Adoutte A. 1995. Genetic code
560 deviations in the ciliates: evidence for multiple and independent events. EMBO J.
561 14:3262–3267.

562 Waskom M. 2021. seaborn: statistical data visualization. JOSS 6:3021.

563 Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva
564 EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene
565 prediction and phylogenomics. Mol. Biol. Evol. 35:543–548.

566 Xiong J, Wang G, Cheng J, Tian M, Pan X, Warren A, Jiang C, Yuan D, Miao W. 2015.
567 Genome of the facultative scuticociliatosis pathogen *Pseudocohnilembus persalinus*
568 provides insight into its virulence through horizontal gene transfer. Sci. Rep. 5:15470.

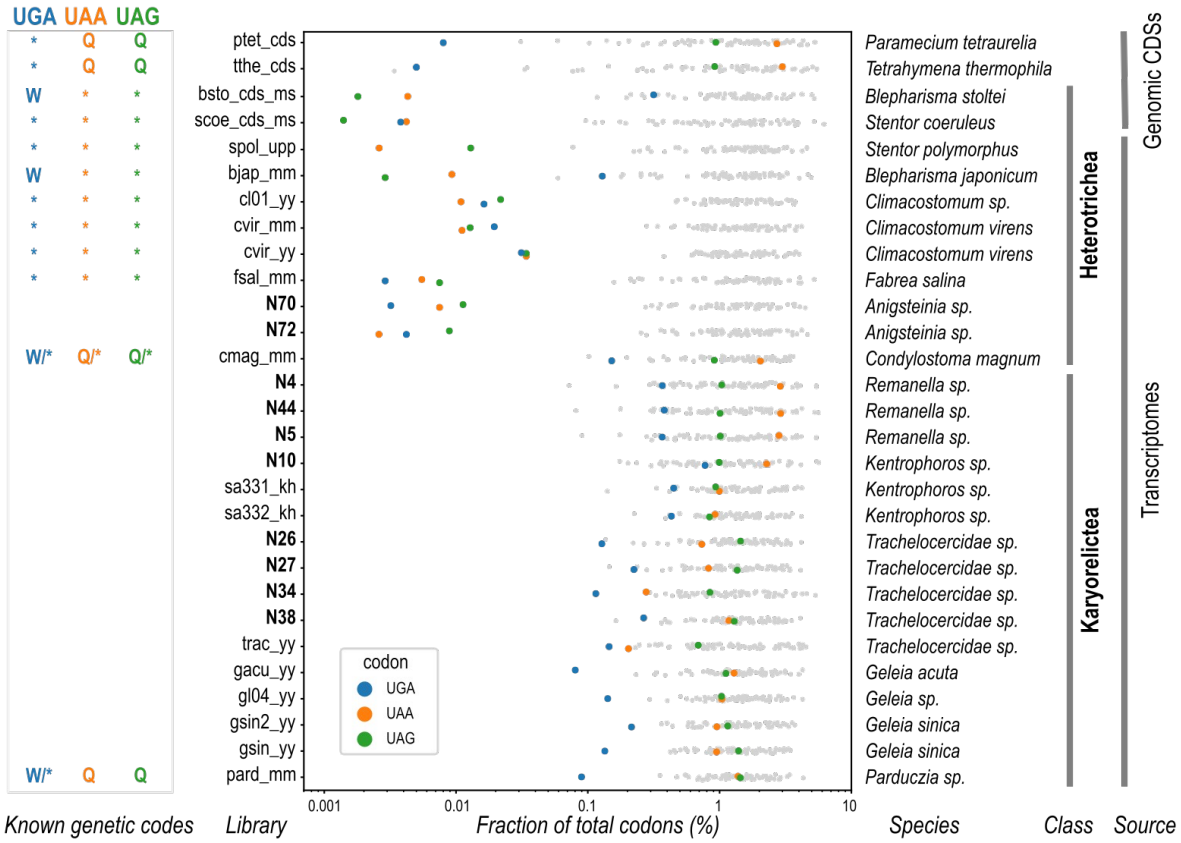
569 Xu Y, Li J, Song W, Warren A. 2013. Phylogeny and establishment of a new ciliate family,
570 Wilbertomorphidae fam. nov. (Ciliophora, Karyorelictea), a highly specialized taxon
571 represented by *Wilbertomorpha colpoda* gen. nov., spec. nov. J. Eukaryot. Microbiol.
572 60:480–489.

573 Yan Y, Maurer-Alcalá XX, Knight R, Kosakovsky Pond SL, Katz LA. 2019. Single-Cell
574 Transcriptomics Reveal a Correlation between Genome Architecture and Gene Family
575 Evolution in Ciliates. MBio 10.

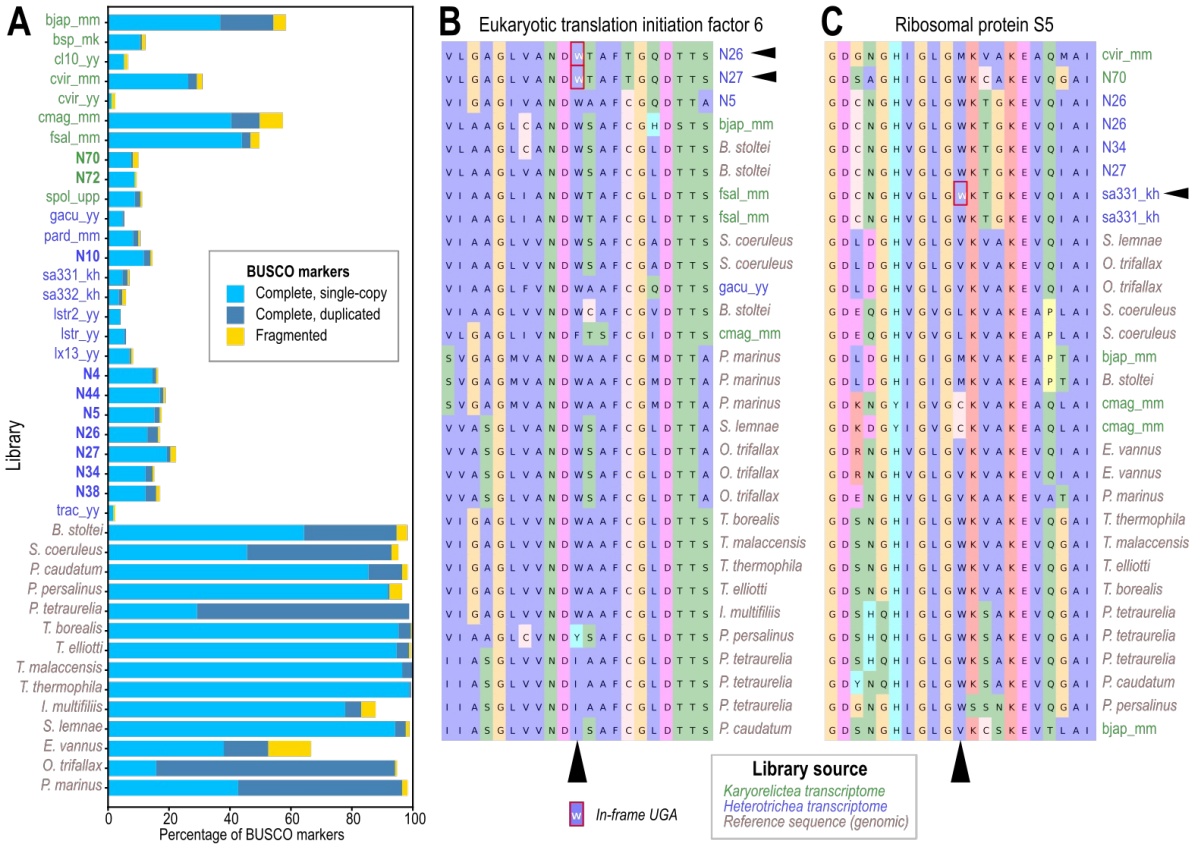
576 Záhonová K, Kostygov AY, Ševčíková T, Yurchenko V, Eliáš M. 2016. An Unprecedented
577 Non-canonical Nuclear Genetic Code with All Three Termination Codons Reassigned as
578 Sense Codons. Curr. Biol. 26:2364–2369.

Figures

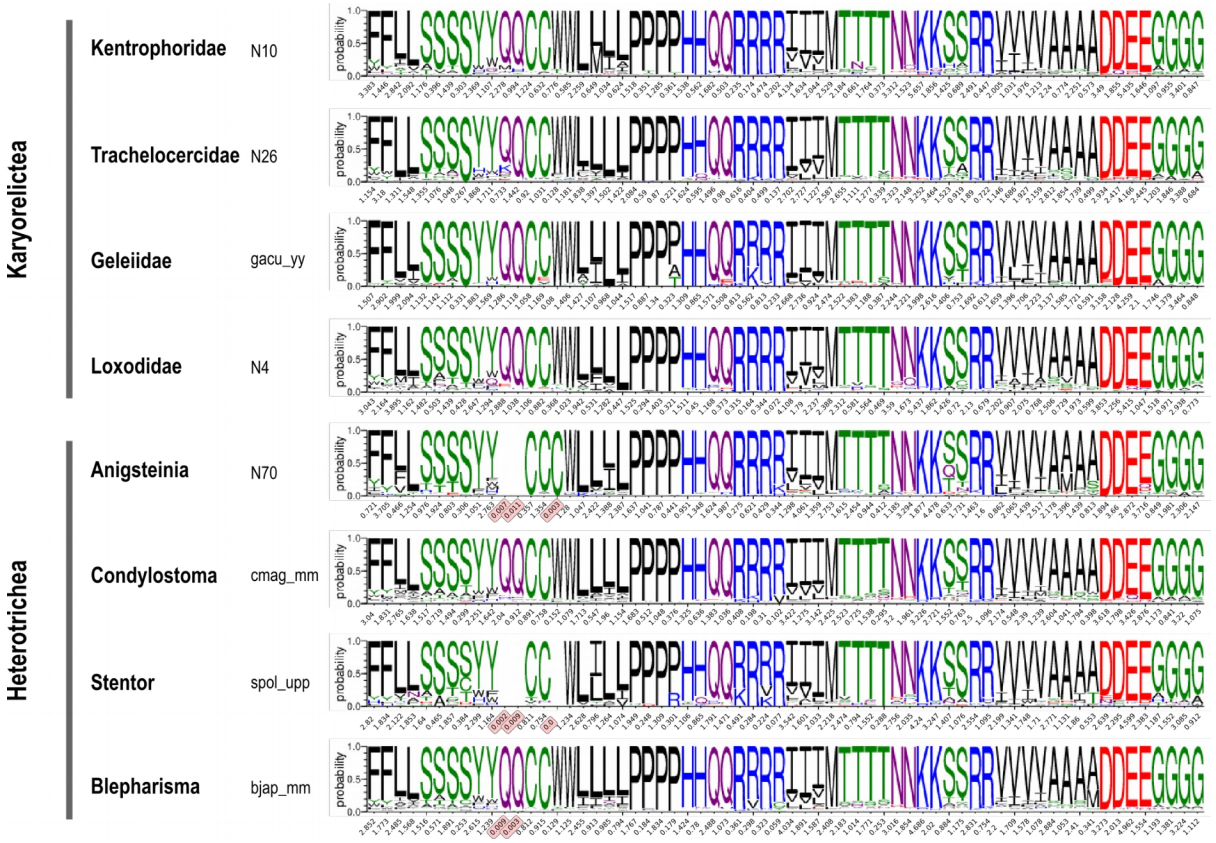
Figure 1.



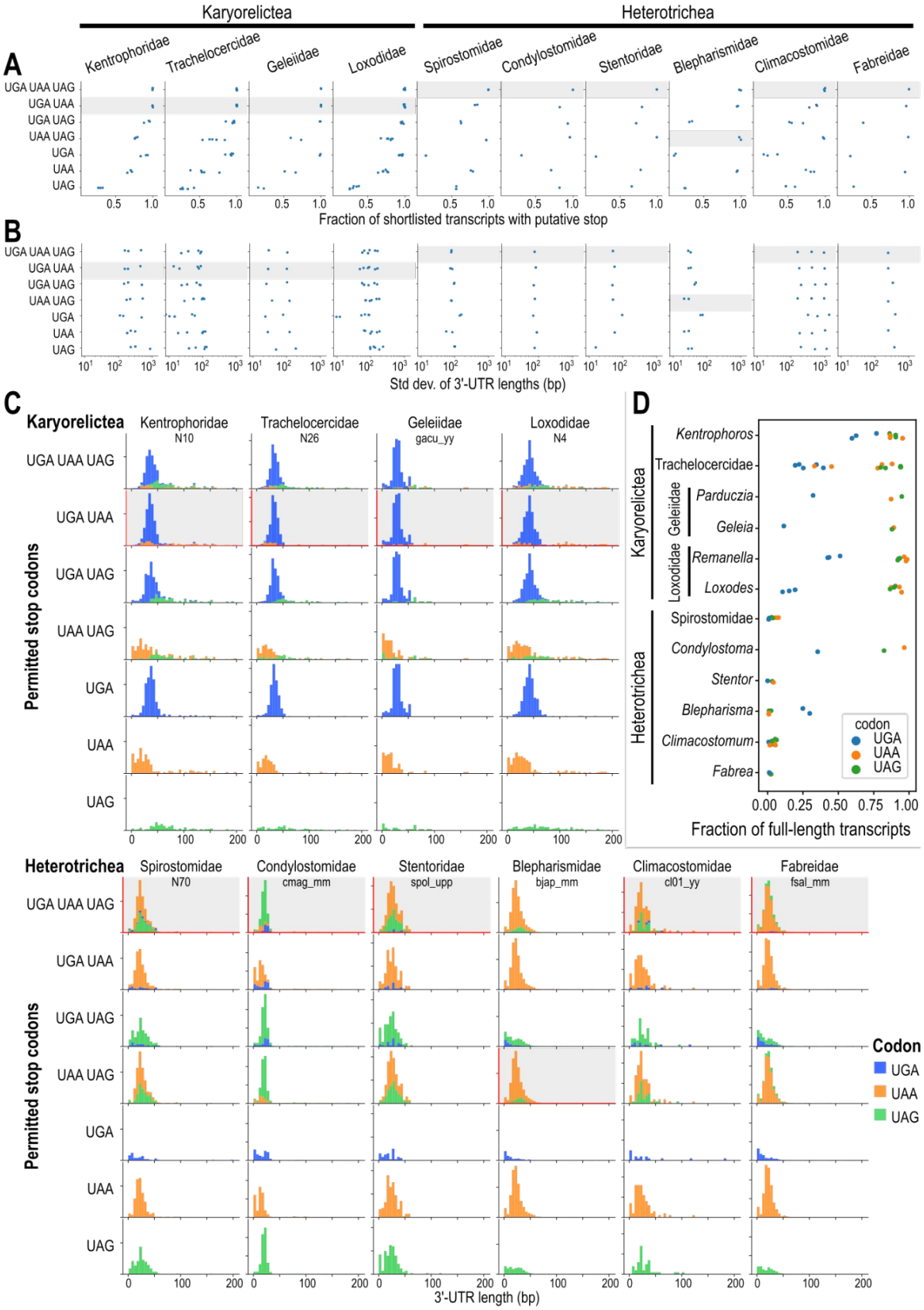
581 *Figure 2.*


A
B
Eukaryotic translation initiation factor 6
C
Ribosomal protein S5
Library
Percentage of BUSCO markers
BUSCO markers
Complete, single-copy
Complete, duplicated
Fragmented
In-frame UGA
Library source
Karyorelictea transcriptome
Heterotrichea transcriptome
Reference sequence (genomic)

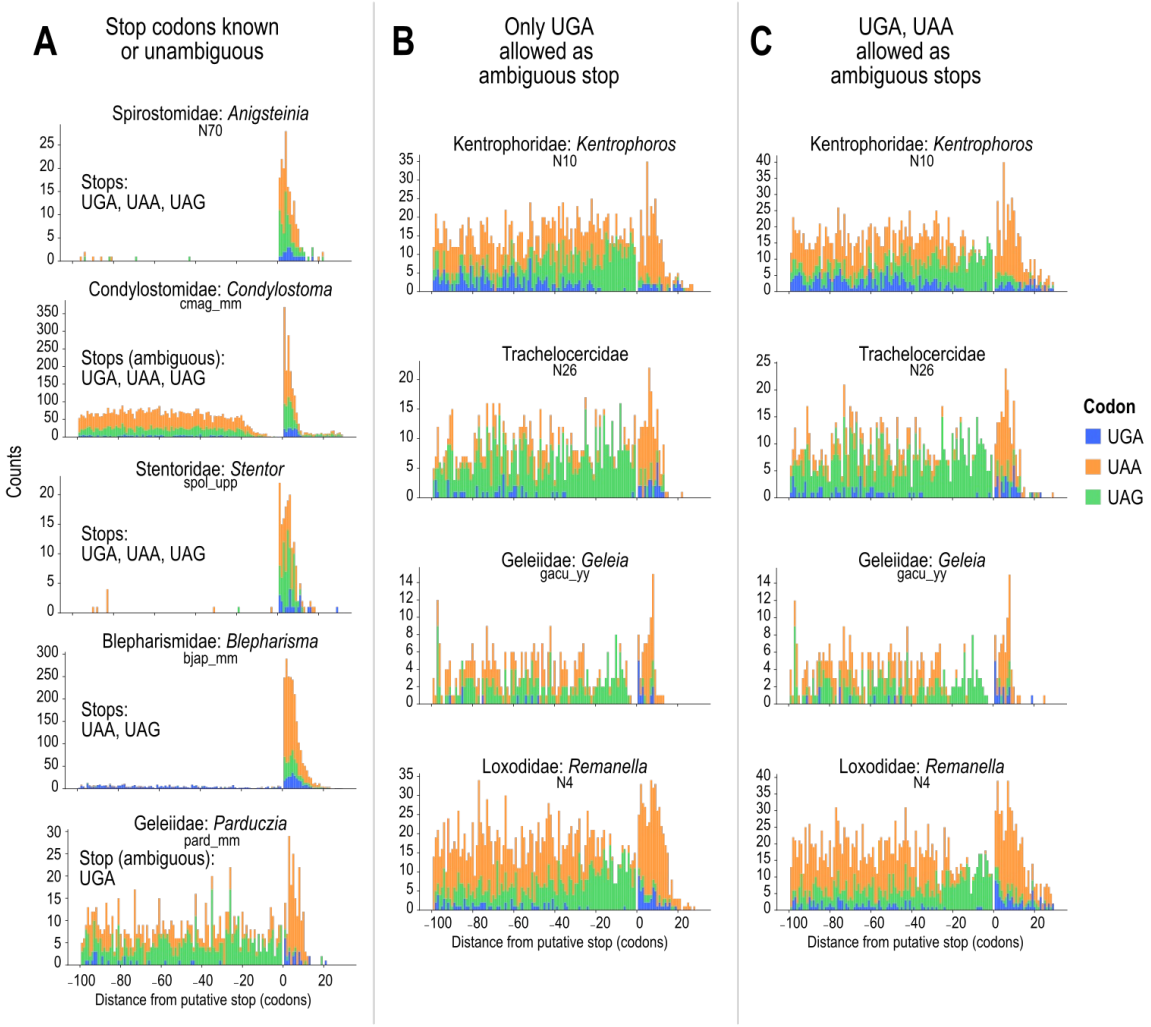
582 *Figure 3.*


Karyorelictea
Heterotrichea
Kentrophoridae N10
Trachelocercidae N26
Geleiidae gacu_yy
Loxodidae N4
Anigsteinia N70
Condylostoma cmag_mm
Stentor spol_upp
Blepharisma bjap_mm
probability

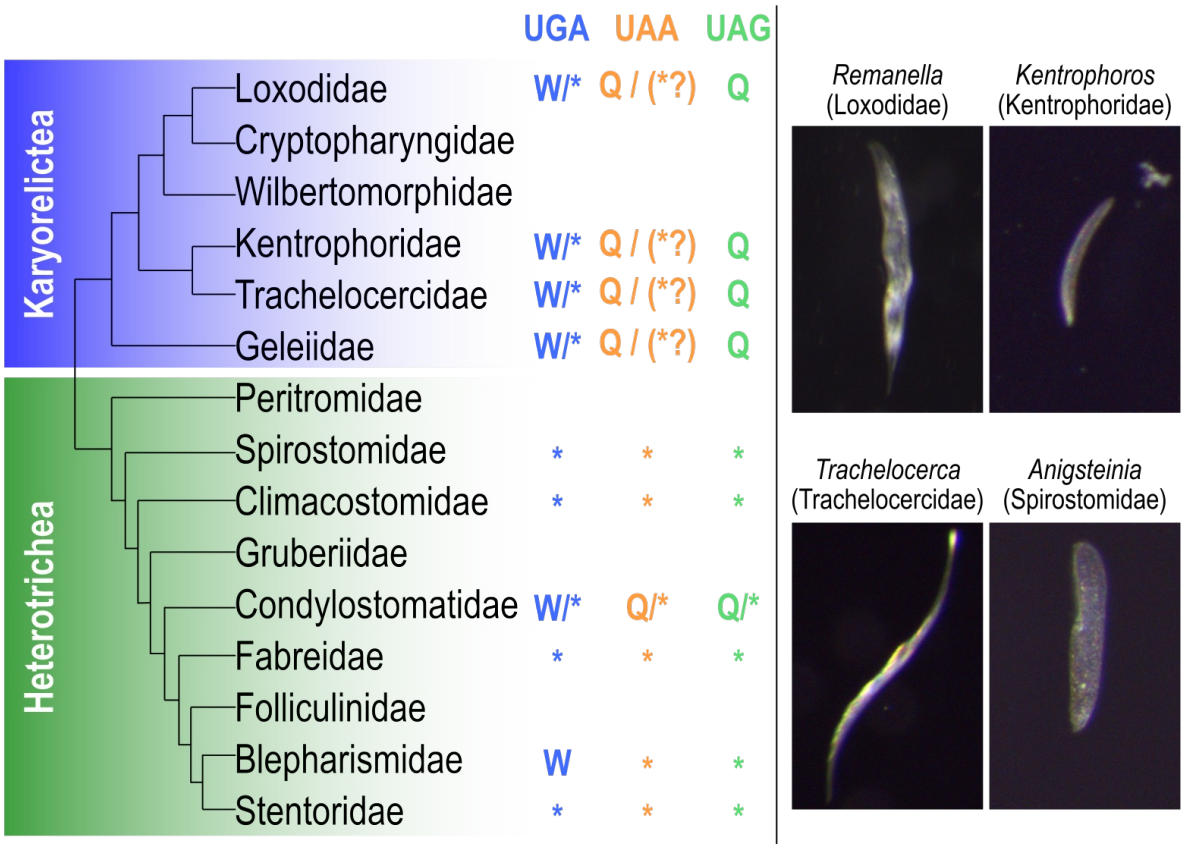
583 *Figure 4.*


Karyorelictea
Heterotrichea
A
Kentrophoridae
Trachelocercidae
Geleiidae
Loxodidae
Spirostomidae
Condylostomidae
Stentoridae
Blepharismidae
Climacostomidae
Fabreidae
UGA UAA UAG
UGA UAA
UGA UAG
UAA UAG
UGA
UAA
UAG
Fraction of shortlisted transcripts with putative stop
B
Std dev. of 3'-UTR lengths (bp)
C
Karyorelictea
Kentrophoridae N10
Trachelocercidae N26
Geleiidae gacu_yy
Loxodidae N4
Permitted stop codons
D
Kentrophoros
Trachelocercidae
Parduczia
Geleia
Remanella
Loxodes
Spirostomidae
Condylostoma
Stentor
Blepharisma
Climacostomum
Fabrea
Karyorelictea
Geleiidae
Loxodidae
Heterotrichea
codon
UGA
UAA
UAG
Fraction of full-length transcripts
Heterotrichea
Spirostomidae N70
Condylostomidae cmag_mm
Stentoridae spol_upp
Blepharismidae bjap_mm
Climacostomidae cl01_yy
Fabreidae fsal_mm
Codon
UGA
UAA
UAG
3'-UTR length (bp)

584 *Figure 5.*


A
Stop codons known or unambiguous
B
Only UGA allowed as ambiguous stop
C
UGA, UAA allowed as ambiguous stops
Spirostomidae: Anigsteinia
N70
Stops: UGA, UAA, UAG
Condylostomidae: Condylostoma
cmag_mm
Stops (ambiguous): UGA, UAA, UAG
Stentoridae: Stentor
spol_upp
Stops: UGA, UAA, UAG
Blepharismidae: Blepharisma
bjap_mm
Stops: UAA, UAG
Geleiidae: Parduczia
pard_mm
Stop (ambiguous): UGA
Counts
Distance from putative stop (codons)
Kentrophoridae: Kentrophoros
N10
Trachelocercidae
N26
Geleiidae: Geleia
gacu_yy
Loxodidae: Remanella
N4
Codon
UGA
UAA
UAG

585 *Figure 6.*


UGA UAA UAG
Karyorelictea
Loxodidae W/* Q / (*?) Q
Cryptopharyngidae
Wilbertomorphidae
Kentrophoridae W/* Q / (*?) Q
Trachelocercidae W/* Q / (*?) Q
Geleiidae W/* Q / (*?) Q
Heterotrichea
Peritromidae
Spirostomidae * * *
Climacostomidae * * *
Gruberiidae
Condylostomatidae W/* Q/* Q/*
Fabreidae * * *
Folliculinidae
Blepharismidae W * *
Stentoridae * * *
Remanella (Loxodidae)
Kentrophoros (Kentrophoridae)
Trachelocerca (Trachelocercidae)
Anigsteinia (Spirostomidae)