Report

Current Biology

An Unprecedented Non-canonical Nuclear Genetic Code with All Three Termination Codons Reassigned as Sense Codons

Highlights

- We discovered a new variant of the nuclear genetic code in a trypanosomatid lineage
- UGA encodes tryptophan; UARs encode glutamate or designate translation termination
- Such a dual meaning of UAR codons is unprecedented

Authors

Kristína Záhonová, Alexei Y. Kostygov, Tereza Ševčíková, Vyacheslav Yurchenko, Marek Eliáš

Correspondence

marek.elias@osu.cz

In Brief

Záhonová et al. demonstrate that a poorly studied lineage of trypanosomatids exhibits a new variant of the nuclear genetic code with all three standard termination codons reassigned to code for amino acids. UAA and UAG at the same time serve as genuine termination codons. This discovery sheds new light on genetic code evolution.

Accession Numbers

KX138599
KX138600
KX138601
KX138602
KX138603

Záhonová et al., 2016, Current Biology *26*, 2364–2369
September 12, 2016
http://dx.doi.org/10.1016/j.cub.2016.06.064

CellPress

Current Biology
Report

An Unprecedented Non-canonical Nuclear Genetic Code with All Three Termination Codons Reassigned as Sense Codons

Kristína Záhonová,[1,2,5] Alexei Y. Kostygov,[1,3,5] Tereza Ševčíková,[1] Vyacheslav Yurchenko,[1,2,4] and Marek Eliáš[1,2,*]
[1]Life Science Research Centre, Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic
[2]Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic
[3]Zoological Institute of the Russian Academy of Sciences, Universitetskaya nab. 1, 199034 Saint Petersburg, Russia
[4]Institute of Parasitology, Czech Academy of Sciences, Branišovská 31, 370 05 České Budějovice, Czech Republic
[5]Co-first author
*Correspondence: marek.elias@osu.cz
http://dx.doi.org/10.1016/j.cub.2016.06.064

SUMMARY

A limited number of non-canonical genetic codes have been described in eukaryotic nuclear genomes. Most involve reassignment of one or two termination codons as sense ones [1–4], but no code variant is known that would have reassigned all three termination codons. Here, we describe such a variant that we discovered in a clade of trypanosomatids comprising nominal *Blastocrithidia* species. In these protists, UGA has been reassigned to encode tryptophan, while UAG and UAA (UAR) have become glutamate encoding. Strikingly, UAA and, less frequently, UAG also serve as bona fide termination codons. The release factor eRF1 in *Blastocrithidia* contains a substitution of a conserved serine residue predicted to decrease its affinity to UGA, which explains why this triplet can be read as a sense codon. However, the molecular basis for the dual interpretation of UAR codons remains elusive. Our findings expand the limits of comprehension of one of the fundamental processes in molecular biology.

RESULTS AND DISCUSSION

While studying the RABL2 gene [5], we noticed that the transcriptome shotgun assembly (TSA) from the bug *Lygus hesperus* [6] surprisingly includes a sequence highly similar to trypanosomatid RABL2 orthologs, yet contains multiple in-frame termination codons. Subsequent searches of the *L. hesperus* TSA revealed additional trypanosomatid-like sequences with the same feature (Supplemental Experimental Procedures; Figure 1A; Figure S1). Importantly, all three termination codons, i.e., UAG, UAA, and UAA, could be found apparently interrupting coding sequences in the trypanosomatid-like contigs, often all three in the same coding sequence. Altogether we examined 125 contigs with obvious trypanosomatid affinity, 100 of which featured at least a partial splice leader (SL) sequence, i.e., an invariant region at the 5′ end of mRNA molecules attached by the process of *trans*-splicing (Table S1) [7]. This indicates that most, if not all, of these contigs correspond to mRNA molecules rather than to contaminating DNA. Of the 125 transcripts investigated, 77 included at least one of the UAG, UAA, or UAA codons within their predicted coding sequences (i.e., upstream of the presumed genuine termination codons; see below). This suggested that the TSA from *L. hesperus* is contaminated by a trypanosomatid-related organism employing a non-canonical genetic code that has reassigned UAG, UAA, and UGA as sense codons.

Meadow bugs (family Miridae), which include *L. hesperus*, are known to host trypanosomatids [8]. Therefore, we assumed that the *L. hesperus* individuals used in the transcriptome sequencing project had been infected with these flagellates. Indeed, we were able to assemble a full trypanosomatid 18S rRNA sequence from the respective raw RNA sequencing (RNA-seq) reads. It proved to be identical to an 18S rRNA gene sequence that we obtained from a trypanosomatid inhabiting a *Lygus* sp. bug collected in Northern Karelia, Russia (Supplemental Experimental Procedures). This parasite, perhaps, represents a hitherto undescribed species and belongs to a clade composed of species assigned to the genus *Blastocrithidia* (Figure 1B). Note that the taxonomy of this trypanosomatid group is poorly studied [8, 9], and its true generic affinity is uncertain. Our unpublished results suggest that the type species of the genus *Blastocrithidia* may represent a separate lineage. Hence, classification of the clade including the parasite of *L. hesperus* as *Blastocrithidia* should be considered provisional.

We searched the *L. hesperus* TSA for homologs of conserved genes used previously in a multi-protein analysis of the trypanosomatid phylogeny [10]. Trypanosomatid orthologs were found for 64 of them, one for each gene (Table S1), suggesting infection by one trypanosomatid species hereafter referred to as *Blastocrithidia* sp. An analysis of kinetoplastid phylogeny using a supermatrix of amino acid sequences deduced for the respective 64 genes (Figure S2A) showed *Blastocrithidia* sp. as an isolated lineage possibly related to the recently proposed subfamily Phytomonadinae [11]. The lack of close relatives of the *Blastocrithidia* sp. in the phylogenomic tree is an expected result given the lack of genome-scale data from *Blastocrithidia* species. However, the assignment of this trypanosomatid to the *Blastocrithidia* clade was validated by close relationship of its HSP83

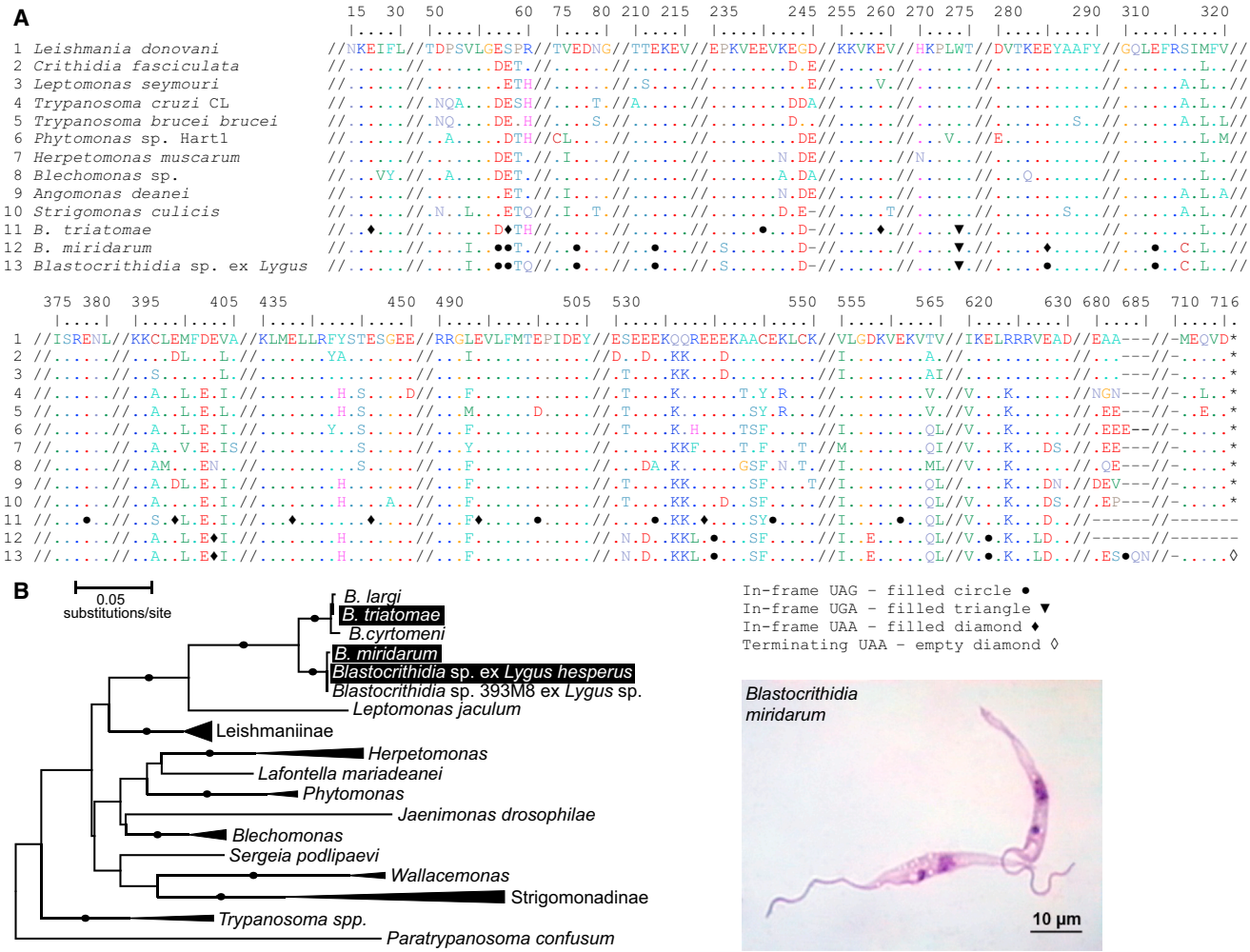

Figure 1. A New Genetic Code in a Clade of Trypanosomatids

(A) Multiple sequence alignment of trypanosomatid HSP83 amino acid sequences. Positions corresponding to UAG, UAA, and UGA codons are indicated by symbols explained in the graphical legend. Slashes represent regions of the sequence alignment (without these codons) that were omitted for simplicity. Dots indicate amino acid identity with the top-most sequence; the numbers above the sequence alignment show the actual position in the full sequence alignment. The sequences from *Blastocrithidia triatomae* and *B. miridarum* are incomplete at the C termini.

(B) A phylogenetic tree of trypanosomatids inferred from 18S rRNA sequences using the maximum likelihood method (details on the phylogenetic analysis are provided in Supplemental Experimental Procedures). Dots mark branches with maximal statistical support. The species with the novel genetic code (see Figure 1A) are highlighted in black. The micrograph of *Blastocrithidia miridarum* is shown courtesy of Dr. A.O. Frolov.

See also Figures S1 and S2 and Table S1.

sequence with those we obtained from two identified *Blastocrithidia* species (*B. miridarum* and *B. triatomae*; Figure S2B). Moreover, HSP83 gene sequences of both species also exhibited in-frame UAG, UAA, and UGA codons (Figure 1A). This indicates that the non-canonical genetic code is inherent to representatives of both main *Blastocrithidia* subclades (Figure 1B), suggesting that this feature was present in an ancestor of the whole clade.

The known non-canonical genetic code variants of eukaryotic nuclear genomes belong to two categories. In the first category, the meaning of a particular codon is changed from one amino acid to another, but such a variant has so far been described only from two groups of yeasts [12, 13]. The second, more common category includes codes changing a standard termination codon to a sense one [1, 3, 4]. Reassignment of UGA has been so far restricted to some ciliates. Specifically, UGA is used as a cysteine codon in *Euplotes* spp. [14] and as a tryptophan codon in at least two independent ciliate lineages, the genus *Colpoda* and a subgroup of the class Heterotrichea (e.g., *Stentor* and *Blepharisma*) [15]. Several other ciliate lineages have retained UGA as a termination codon and reassigned UAR as sense codons specifying glutamate or glutamine. The former variant was reported from peritrich ciliates *Vorticella* and *Opisthonecta* [2]. The usage of UAR for glutamine appears to have evolved in several ciliate lineages independently (from four to six independent origins of this genetic code variant in ciliates have been proposed; [1, 15]). Moreover, UAR reassignment to glutamine has been reported from a growing number of

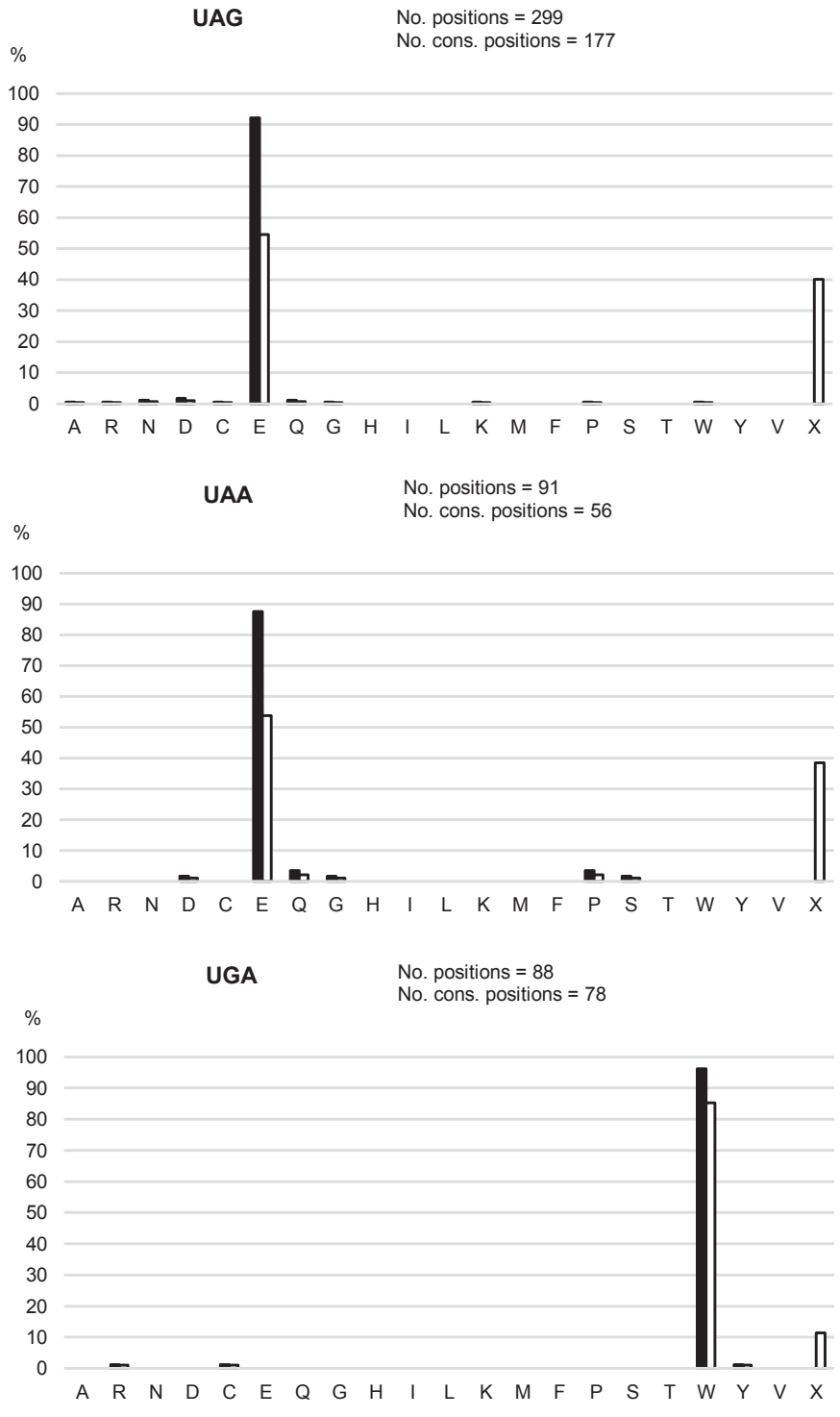

Figure 2. In-Frame UAR and UGA Codons in the Genes from *Blastocrithidia* sp. Typically Correspond to Positions with a Conserved Glutamate and Tryptophan Residue, Respectively

Fifty-five proteins encoded by the putative *Blastocrithidia* sp. transcripts were compared to their homologs from other kinetoplastids, and positions in the sequence alignment of the homologous sequences occupied by in-frame UAG, UAA, and UGA codons were analyzed. The one-letter amino acid abbreviations along the x axis indicate positions with a particular amino acid residue conserved in more than 75% sequences; "X" represents positions where no amino acid residue reached such a degree of conservation. The y axis indicates the percentage of the different positions. White bars indicate percentage calculated from all positions occupied by the given non-standard codon; black bars indicate percentage calculated only for conserved positions (i.e., excluding position in the category "X"). See also Figure S1 and Table S1.

eukaryotic taxa outside ciliates. Those include hexamitid diplomonads [16], a subset of ulvophytes [17], some oxymonads [18], and *Amoebaphelidium protococcarum*, a member of the Aphelidea (algal parasites related to Fungi) [19].

Despite the expanding list of eukaryotes with deviating genetic codes for nuclear genes, the presumed reassignment of all three termination codons as sense ones we revealed in the *Blastocrithidia* sp. transcripts is unprecedented to our knowledge (but see Note Added in Proof below). Therefore, we investigated features of this apparently novel code variant in detail. First, we tried to establish the meaning of the reassigned UAG, UAA, and UGA codons. To this end, we analyzed the multiple sequence alignments of the 64 *Blastocrithidia* sp. sequences and their kinetoplastid orthologs that we used in the phylogenomic analysis above. Of these, 55 contained at least one apparent in-frame termination codon (UAG, UAA, or UGA) in sequences from *Blastocrithidia* sp. (examples provided in Figure S1). We investigated the identity and the degree of amino acid conservation at positions corresponding to these in-frame termination codons. Each position was considered as conserved if a given amino acid was present in at least 75% of the sequences in the sequence alignment; the remaining positions were treated as non-conserved. 92.09% of conserved positions with in-frame UAG and 87.50% of conserved positions with in-frame UAA corresponded to glutamate, whereas the percentage of conserved positions with any other particular amino acid did not exceed 4% (Figure 2). Among the conserved positions with in-frame UGA, 96.15% corresponded to tryptophan. When both conserved and non-conserved positions were considered, the above values decreased to 54.52% for UAG, 53.58% for UAA, and 85.23% for UGA positions (Figure 2).

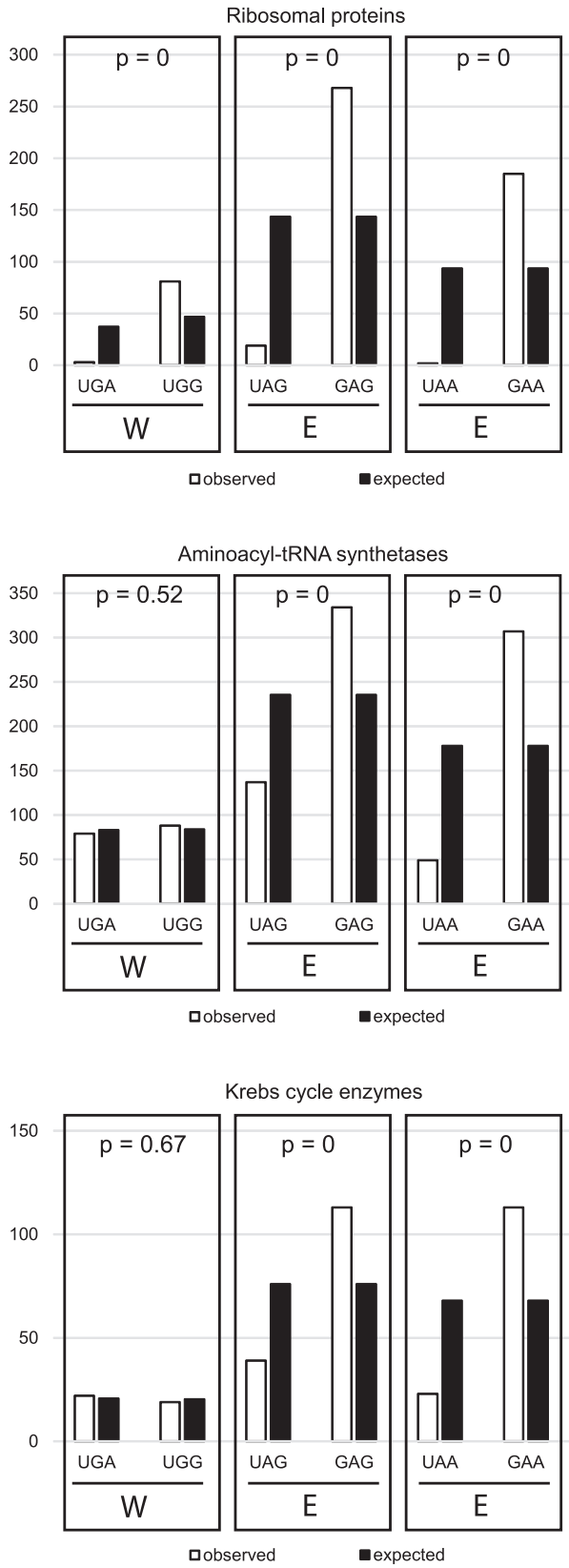

Figure 3. Biases in the Usage of Standard and Non-standard Synonymous Codons in *Blastocrithidia* sp.

Observed and expected (i.e., theoretical unbiased) numbers of the codons were calculated for three different gene sets as described in Supplemental Experimental Procedures. The differences between the observed and expected numbers were tested using the chi-square test (separately for the UGA–UGG, UAG–GAG, and UAA–GAA pairs). All differences were statistically significant, except the differences between the observed and expected numbers concerning the UGA–UGG pairs in aminoacyl-tRNA synthetases and Krebs cycle enzymes ($p > 0.05$ in both cases).

We assume that each codon is translated as the same amino acid at all positions (there is only one known exception to this rule: the ciliate *Euplotes crassus* translating UGA as cysteine in most positions or as selenocysteine in certain positions defined by the SECIS element [20]). Hence, the analysis above suggests that in *Blastocrithidia* sp., UAG and UAA are translated as glutamate and UGA is translated as tryptophan. None of these reassignments is without a precedent. As discussed above, UAA and UAG encoding glutamate are known from nuclear genes of the ciliates *Vorticella* and *Opisthonecta*, but UGA in these species apparently serves as a termination codon and does not encode tryptophan [2]. On the other hand, UGA reassigned as a tryptophan codon is one of the most common deviations from the standard genetic code. It is known from some bacterial, mitochondrial, and plastid translation systems [1], as well as from the nuclear genomes of two lineages of ciliates (different from that exhibiting UAR as glutamate codons [15]).

Strikingly, UAG and UAA are used not only as glutamate codons in *Blastocrithidia* sp., but both apparently still designate translation termination (Figure 1A; Figure S1). Based on conservation of C-terminal sequences of trypanosomatid orthologs, we could define putative bona fide termination codons for 115 out of 125 *Blastocrithidia* sp. transcripts analyzed (the remaining ten transcript sequences were either incomplete at the 3′ end or there were multiple equally likely candidates for the termination codon in them). In this sample of inferred termination codons, UAA predominated (109 cases), while UAG was used rarely (only in six transcripts), and no coding sequence was found to be terminated with UGA (Table S1). In contrast, UAG was used more often as a sense codon than UAA in the same sample of 115 transcripts (404 versus 124 instances, respectively).

Next, we investigated to what extent the non-standard codons have already become integrated into the genetic vocabulary of *Blastocrithidia* sp. by checking possible bias in usage of the non-standard codons compared to synonymous standard codons for glutamate and tryptophan. We specifically focused on three sets of genes representing different functional categories: (1) ribosomal proteins (55 genes), (2) aminoacyl-tRNA synthetases (15 genes), and (3) enzymes of the Krebs cycle (8 genes) (Table S1). We then calculated expected numbers of each of the glutamate (UAG, UAA, GAG, GAA) and tryptophan (UGA and UGG) codons separately for each gene category and compared them with the actual observed numbers (Supplemental Experimental Procedures). We revealed that genes for ribosomal proteins use UAG, UAA, and UGA codons with a significantly lower frequency than expected (Figure 3). In the remaining two gene categories, UAG and especially UAA are also

significantly depleted, but UGA is used with the same frequency as UGG (Figure 3).

Two factors, not necessarily mutually exclusive, need to be considered as possible causes of the observed bias. First, the depletion of the non-standard codons as compared to the standard synonymous ones (only in some gene categories in case of UGA versus UGG) may reflect the fact that a certain evolutionary time is needed for the newly reassigned non-standard codons to spread across the genome and equilibrate in frequency with the originally present synonymous standard codons. However, the phylogenetically broad occurrence of the new code variant in the *Blastocrithidia* clade suggests that the change in the code is not a particularly recent event. Hence, negative selection acting against the spreading of the non-standard codons in coding sequences of *Blastocrithidia* may be the actual factor behind the pattern observed. This is apparent from the fact that the different gene categories exhibit a different degree of depletion of the three non-standard codons (Figure 3). UGA is not discriminated at all in some gene categories (aminoacyl-tRNA synthetases, Krebs cycle enzymes), indicating that its frequency has equilibrated with UGG. However, it is strongly depleted in genes for ribosomal proteins, which is a gene category known to require highly efficient translation of the respective mRNAs, imposing a particularly strong pressure on codon usage [21]. Assuming that the main entry of both UAG and UAA codons into the *Blastocrithidia* coding sequences is the same G-to-U transversion of the first nucleotide in GAG and GAA codons, respectively, the selectively neutral model predicts that the UAG-to-GAG and UAA-to-GAA ratios should be roughly the same. However, the latter ratio actually is substantially smaller in all three gene categories tested (Figure 3). Hence, there may be stronger negative selection operating against in-frame UAA codons, apparently because UAA is the main termination codon in the *Blastocrithidia* genetic code.

The efficiency of different synonymous codons in translation is primarily dictated by the relative abundances of the corresponding isoacceptor aminoacyl-tRNA molecules [4, 22]. Therefore, it is possible that the observed bias against the usage of UAG, UAA, and UGA codons is at least partly due to limited abundance of hypothetical cognate aminoacyl-tRNAs. However, efficient accommodation of any of UAG, UAA, or UGA codons as sense ones requires not only the existence of (sufficiently abundant) cognate aminoacyl-tRNAs but also specific modifications in the eukaryotic release factor 1, eRF1 [23]. This protein is responsible for recognizing all three termination codons in an mRNA during translation and liberating the nascent polypeptide chain from the tRNA that has brought the last (C-terminal) amino acid residue [24]. Reassignment of any of the termination codons thus necessitates narrowing the specificity of the eRF1 protein to avoid interference with the proper function of the reassigned codon. Indeed, studies of eRF1 sequences in eukaryotes with such reassignments revealed that the altered specificity is accompanied by changes in highly conserved regions of the protein [23]. There are several known conserved elements in eRF1 important for termination codon recognition: GTS, Glu55, (TAS) NIKS, and YxCxxxF [25]. Eukaryotes that use non-standard genetic codes typically exhibit alternations in these motifs [23, 26]. Intriguingly, eRF1 in *Blastocrithidia* sp. shows no changes in any of these standard motifs (Figure S3).

In addition, the eRF1 proteins of unrelated ciliate lineages with non-canonical genetic codes were shown to possess several unique convergently acquired amino acid substitutions [26]. Specifically, ciliates utilizing UGA as a sense codon exhibit a characteristic Ser70Ala mutation. Indeed, an alanine scanning mutagenesis of the yeast eRF1 demonstrated that the Ser70 residue is critical for recognition of the UGA codon by the eRF1 protein [27]. Therefore, it seems significant that the eRF1 protein from *Blastocrithidia* sp. exhibits an equivalent Ser70Gly substitution, which is not shared with other investigated trypanosomatid species employing the standard genetic code (Figure S3). Hence, the *Blastocrithidia* sp. eRF1 is predicted to ignore UGA as a termination codon, although it remains to be tested whether the affinity to this codon is completely lost or just decreased. The latter possibility is suggested by the bias against UGA in genes for ribosomal proteins of *Blastocrithidia* sp. (see above). However, it is presently unclear what might be the modifications of eRF1 that allow *Blastocrithidia* sp. to use UAG and UAA as sense codons and how the eRF1 protein discriminates individual UAR sense codons from bona fide termination codons. The negative bias in UAR (especially UAA) usage may indicate that this discrimination is inefficient. Indeed, such a situation is predicted to occur under the "ambiguous intermediate" mechanism of genetic code evolution [1, 28].

In summary, we provide the first evidence for the presence of a non-canonical nuclear genetic code in trypanosomatids, a well-studied group comprising some deadly human pathogens. Moreover, this genetic code is unprecedented, and its features point to an unanticipated functional and evolutionary flexibility of the translation system. Detailed genomic and biochemical analyses are needed to unveil the actual molecular mechanism behind the peculiar genetic code in *Blastocrithidia*.

ACCESSION NUMBERS

The accession numbers for the DNA sequences reported in this paper are GenBank: KX138599, KX138600, KX138601, KX138602, and KX138603.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.06.064.

AUTHOR CONTRIBUTIONS

K.Z. and A.Y.K. contributed equally to this work. K.Z. performed most of the analyses related to the new code and the phylogenomic analysis and prepared most of the figures and Table S1. A.Y.K. obtained the new sequences from *Blastocrithidia* sp. isolates and performed phylogenetic analyses of 18S rRNA and HSP83. T.Ś. contributed to the sequence analyses and prepared Figure S3. V.Y. contributed to the design of the whole project and the analyses. M.E. discovered the trypanosomatid contamination in the *Lygus hesperus* transcriptome assembly, conceived the project, and wrote the paper with input from all authors.

ACKNOWLEDGMENTS

We thank Dr. Alexander O. Frolov for providing the image of *Blastocrithidia miridarum* and two anonymous reviewers for constructive comments on the original version of the paper. This study was supported by Czech Science Foundation grants 13-24983S (to M.E.) and 16-18699S (to V.Y.) and by the

project LO1208 of the National Feasibility Programme I of the Czech Republic. A.Y.K. was supported by grant no. 0924/2016/ŠaS from the Statutory City of Ostrava.

Received: May 15, 2016
Revised: June 25, 2016
Accepted: June 28, 2016
Published: September 1, 2016

REFERENCES

1. Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat. Rev. Genet. *2*, 49–58.
2. Sánchez-Silva, R., Villalobo, E., Morin, L., and Torres, A. (2003). A new noncanonical nuclear genetic code: translation of UAA into glutamate. Curr. Biol. *13*, 442–447.
3. Ohama, T., Inagaki, Y., Bessho, Y., and Osawa, S. (2008). Evolving genetic code. Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. *84*, 58–74.
4. Ling, J., O'Donoghue, P., and Söll, D. (2015). Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. Nat. Rev. Microbiol. *13*, 707–721.
5. Eliáš, M., Klimeš, V., Derelle, R., Petrželková, R., and Tachezy, J. (2016). A paneukaryotic genomic analysis of the small GTPase RABL2 underscores the significance of recurrent gene loss in eukaryote evolution. Biol. Direct *11*, 5.
6. Hull, J.J., Chaney, K., Geib, S.M., Fabrick, J.A., Brent, C.S., Walsh, D., and Lavine, L.C. (2014). Transcriptome-based identification of ABC transporters in the western tarnished plant bug *Lygus hesperus*. PLoS ONE *9*, e113046.
7. Liang, X.H., Haritan, A., Uliel, S., and Michaeli, S. (2003). trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot. Cell *2*, 830–840.
8. Maslov, D.A., Yurchenko, V.Y., Jirků, M., and Lukeš, J. (2010). Two new species of trypanosomatid parasites isolated from Heteroptera in Costa Rica. J. Eukaryot. Microbiol. *57*, 177–188.
9. Kostygov, A.Iu., and Frolov, A.O. (2007). [Leptomonas jaculum (Leger, 1902) Woodcock 1914: a leptomonas or a blastocrithidia?]. Parazitologia *41*, 126–136.
10. Flegontov, P., Votýpka, J., Skalický, T., Logacheva, M.D., Penin, A.A., Tanifuji, G., Onodera, N.T., Kondrashov, A.S., Volf, P., Archibald, J.M., and Lukeš, J. (2013). *Paratrypanosoma* is a novel early-branching trypanosomatid. Curr. Biol. *23*, 1787–1793.
11. Yurchenko, V., Kostygov, A., Havlová, J., Grybchuk-Ieremenko, A., Ševčíková, T., Lukeš, J., Ševčík, J., and Votýpka, J. (2016). Diversity of trypanosomatids in cockroaches and the description of *Herpetomonas tarakana* sp. n. J. Eukaryot. Microbiol. *63*, 198–209.
12. Santos, M.A., Ueda, T., Watanabe, K., and Tuite, M.F. (1997). The non-standard genetic code of *Candida* spp.: an evolving genetic code or a novel mechanism for adaptation? Mol. Microbiol. *26*, 423–431.
13. Mühlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H., and Kollmar, M. (2016). A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. Genome Res. *26*, 945–955.
14. Tourancheau, A.B., Tsao, N., Klobutcher, L.A., Pearlman, R.E., and Adoutte, A. (1995). Genetic code deviations in the ciliates: evidence for multiple and independent events. EMBO J. *14*, 3262–3267.
15. Lozupone, C.A., Knight, R.D., and Landweber, L.F. (2001). The molecular basis of nuclear genetic code change in ciliates. Curr. Biol. *11*, 65–74.
16. Keeling, P.J., and Doolittle, W.F. (1997). Widespread and ancient distribution of a noncanonical genetic code in diplomonads. Mol. Biol. Evol. *14*, 895–901.
17. Cocquyt, E., Gile, G.H., Leliaert, F., Verbruggen, H., Keeling, P.J., and De Clerck, O. (2010). Complex phylogenetic distribution of a non-canonical genetic code in green algae. BMC Evol. Biol. *10*, 327.
18. de Koning, A.P., Noble, G.P., Heiss, A.A., Wong, J., and Keeling, P.J. (2008). Environmental PCR survey to determine the distribution of a non-canonical genetic code in uncultivable oxymonads. Environ. Microbiol. *10*, 65–74.
19. Karpov, S.A., Mikhailov, K.V., Mirzaeva, G.S., Mirabdullaev, I.M., Mamkaeva, K.A., Titova, N.N., and Aleoshin, V.V. (2013). Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi. Protist *164*, 195–205.
20. Turanov, A.A., Lobanov, A.V., Fomenko, D.E., Morrison, H.G., Sogin, M.L., Klobutcher, L.A., Hatfield, D.L., and Gladyshev, V.N. (2009). Genetic code supports targeted insertion of two amino acids by one codon. Science *323*, 259–261.
21. Klumpp, S., Dong, J., and Hwa, T. (2012). On ribosome load, codon bias and protein abundance. PLoS ONE *7*, e48542.
22. Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet. *28*, 574–581.
23. Bezerra, A.R., Guimarães, A.R., and Santos, M.A. (2015). Non-standard genetic codes define new concepts for protein engineering. Life (Basel) *5*, 1610–1628.
24. Nakamura, Y., and Ito, K. (1998). How protein reads the stop codon and terminates translation. Genes Cells *3*, 265–278.
25. Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. Nature *524*, 493–496.
26. Inagaki, Y., Blouin, C., Doolittle, W.F., and Roger, A.J. (2002). Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. Nucleic Acids Res. *30*, 532–544.
27. Blanchet, S., Rowe, M., Von der Haar, T., Fabret, C., Demais, S., Howard, M.J., and Namy, O. (2015). New insights into stop codon recognition by eRF1. Nucleic Acids Res. *43*, 3298–3308.
28. Sengupta, S., and Higgs, P.G. (2015). Pathways of genetic code evolution in ancient and modern organisms. J. Mol. Evol. *80*, 229–243.

Note Added in Proof

After this paper had been accepted for publication, two studies, by Swart et al. and Heaphy et al., were published demonstrating the existence of genetic code variants similar to that described by us in *Blastocrithidia*. The ciliate *Condylostoma magnum* proved to use UAR and UGA not only as termination codons but also as sense ones to encode glutamine and tryptophan, respectively, depending on their location within mRNA (Swart et al. and Heaphy et al.). In an unrelated ciliate, *Parduczia* sp., Swart et al. found that UARs serve as glutamine codons, whereas UGA is used dually as a termination or tryptophan codon. These findings are exciting but perhaps less surprising than our report concerning *Blastocrithidia*, given the long-known propensity of ciliates to depart in different ways from the standard genetic code [15].

Swart, E.C., Serra, V., Petroni, G., and Nowacki, M. (2016). Genetic codes with no dedicated stop codon: context-dependent translation termination. Cell *166*, 691–702.

Heaphy, S.M., Mariotti, M., Gladyshev, V.N., Atkins, J.F., and Baranov, P.V. (2016). Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in *C. magnum*. Mol. Biol. Evol. Published online August 8, 2016. 10.1093/molbev/msw166.