Article

Short tRNA anticodon stem and mutant eRF1 allow stop codon reassignment

https://doi.org/10.1038/s41586-022-05584-2

Received: 14 January 2022

Accepted: 18 November 2022

Published online: 11 January 2023

Check for updates

Ambar Kachale[1,2,9], Zuzana Pavlíková[3,9], Anna Nenarokova[1,2,7,9], Adriana Roithová[3], Ignacio M. Durante[1], Petra Miletínová[3], Kristína Záhonová[1,4,5], Serafim Nenarokov[1,2], Jan Votýpka[1,4], Eva Horáková[1,8], Robert L. Ross[6], Vyacheslav Yurchenko[5], Petra Beznosková[3], Zdeněk Paris[1,2]✉, Leoš Shivaya Valášek[3]✉ & Julius Lukeš[1,2]✉

Cognate tRNAs deliver specific amino acids to translating ribosomes according to the standard genetic code, and three codons with no cognate tRNAs serve as stop codons. Some protists have reassigned all stop codons as sense codons, neglecting this fundamental principle[1–4]. Here we analyse the in-frame stop codons in 7,259 predicted protein-coding genes of a previously undescribed trypanosomatid, *Blastocrithidia nonstop*. We reveal that in this species in-frame stop codons are underrepresented in genes expressed at high levels and that UAA serves as the only termination codon. Whereas new $tRNAs^{Glu}$ fully cognate to UAG and UAA evolved to reassign these stop codons, the UGA reassignment followed a different path through shortening the anticodon stem of $tRNA^{Trp}_{CCA}$ from five to four base pairs (bp). The canonical 5-bp $tRNA^{Trp}$ recognizes UGG as dictated by the genetic code, whereas its shortened 4-bp variant incorporates tryptophan also into in-frame UGA. Mimicking this evolutionary twist by engineering both variants from *B. nonstop*, *Trypanosoma brucei* and *Saccharomyces cerevisiae* and expressing them in the last two species, we recorded a significantly higher readthrough for all 4-bp variants. Furthermore, a gene encoding *B. nonstop* release factor 1 acquired a mutation that specifically restricts UGA recognition, robustly potentiating the UGA reassignment. Virtually the same strategy has been adopted by the ciliate *Condylostoma magnum*. Hence, we describe a previously unknown, universal mechanism that has been exploited in unrelated eukaryotes with reassigned stop codons.

The standard genetic code contains 64 codons, 61 of which code for amino acids and three define stop codons, denoting the end of translation[5]. Up to now, at least 30 different alterations of the genetic code are known, most of which are confined to the mitochondrial and bacterial genomes[6–8] (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi). Three types of codon reassignment exist—namely, stop-to-sense, sense-to-sense and sense-to-stop, of which the first category is the most common[9,10], with UAA and UAG reassignments (to glutamine) in the nuclear genomes and UGA in the mitochondrial genomes being the most frequent[3,11,12]. Mechanistically, codon reassignment may involve changes to tRNA anticodons, or tRNA wobble nucleotide modifications, aminoacyl-tRNA synthetase recognition of cognate tRNAs or the fidelity of stop codon recognition by eukaryotic release factors (eRFs)[8]. The most diverse organisms in all of these categories are ciliates harbouring several alternative genetic codes, with stop codon reassignments of various kinds that evolved multiple times independently[13,14].

Several hypotheses explaining evolutionary forces triggering codon reassignments exist. The codon capture model postulates the role of genetic drift in the distribution of specific codons, some of which may become extinct by pressures unrelated to the reassignment, which minimizes its impact on translation[15]. Once they reappear, either owing to altered pressure on the genome or just by chance, they become 'captured' by a near-cognate tRNA of another codon. The alternative 'ambiguous intermediate' theory proposes that spontaneous tRNA mutations change its specificity to recognize also other codons than just the cognate one[16]. This intermediate step is then followed by translational pressure, which induces codon substitutions at positions where ambiguity is deleterious. The 'tRNA loss-driven reassignment' theory proposes an intermediate stage, in which a given codon cannot be translated owing to tRNA gene loss or mutation, creating pressure for synonymous substitutions differing from the original codon to enable its capture by a different tRNA later on[17].

Kinetoplastid protists such as trypanosomes, leishmanias and phytomonads are responsible for diseases of insects, plants and vertebrates, including humans[18]. Although they are known for a wide range of oddities[19], no deviation from the standard genetic code was known from their nuclear genomes until recently. Unexpectedly, a trypanosomatid, *Blastocrithidia* sp., was predicted in silico to reassign all three

[1]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic. [2]Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic. [3]Institute of Microbiology, Czech Academy of Sciences, Prague, Czech Republic. [4]Faculty of Science, Charles University, BIOCEV, Prague, Czech Republic. [5]Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic. [6]Thermo Fisher Scientific, Franklin, MA, USA. [7]Present address: School of Biological Sciences, University of Bristol, Bristol, UK. [8]Present address: Institute of Microbiology, Czech Academy of Sciences, Třeboň, Czech Republic. [9]These authors contributed equally: Ambar Kachale, Zuzana Pavlíková, Anna Nenarokova. ✉e-mail: parda@paru.cas.cz; valasekl@biomed.cas.cz; jula@paru.cas.cz

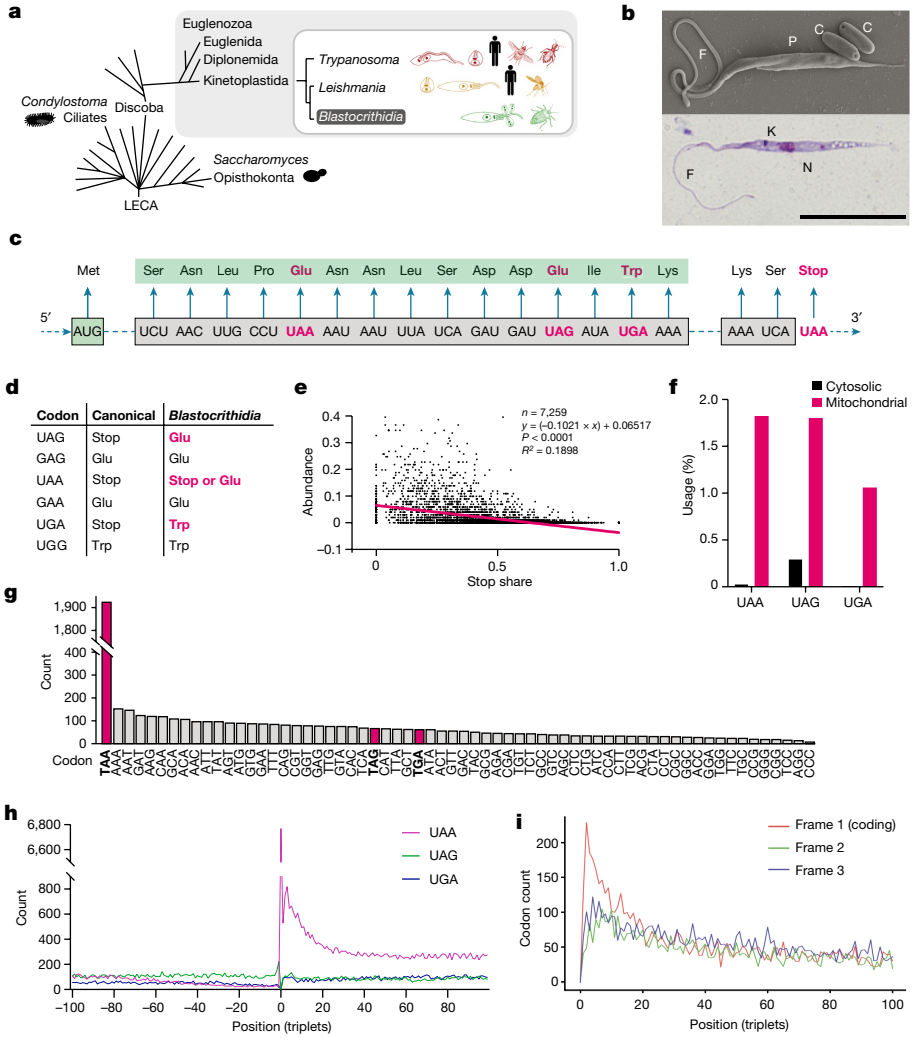


Fig. 1 | Phylogenetic position, morphology, genetic code changes and across-the-genome distribution of reassigned codons in *B. nonstop*. **a**, Phylogenetic tree highlighting the relatedness of insect parasitic *Blastocrithidia* with human parasitic *Trypanosoma* and *Leishmania* (main morphotypes, hosts and vectors are shown by pictograms). LECA, last eukaryotic common ancestor. **b**, Morphology of *B. nonstop*. Scanning electron and light microscopy of cultured cells. K, kinetoplast; N, nucleus; F, flagellum. The promastigote stage (P) dominates over the cyst-like straphanger stage (C); repeated three times. Scale bar, 10 µm. **c**, An example of the coding sequence (homologue of *Tb927.7.4700* encoding a conserved hypothetical protein in *T. brucei*) containing reassigned and genuine stop codons (magenta). **d**, Differences between the canonical and *B. nonstop* genetic codes. **e**, Scatter plot showing non-random distribution of in-frame stop codons. Each of 7,259 evaluated proteins is represented by a dot. The *x*-axis unit shows the relative frequency of reassigned stop codons (stop share). The *y*-axis unit shows the mass-spectrometry-based relative abundance of proteins. The trend line was generated by simple linear regression (Prism). Squared correlation coefficient ($R^2 = 0.1898$) corresponds to two-sided $P < 0.0001$; no multiple comparisons were made. **f**, Abundance of in-frame stop codons in cytosolic (black bars) and mitochondrial (magenta bars) ribosomal proteins of *B. nonstop*. **g**, Codon abundance after the ultimate residue of the reference protein alignments. **h**, Stop codon distribution in the 3′ ends of 7,259 genes, with TAA underrepresented before and overrepresented after the genuine stop codon. **i**, Distribution of UAA in 3′ UTRs of 1,569 *B. nonstop* transcripts. The graph shows the summary counts of UAA in 100 triplets after stop codon calculated in three coding frames; frame 1 corresponds to the gene-coding frame. Equiprobable distribution of UAA in all three frames in the first 20 codons as a null hypothesis was rejected with the two-tailed $P < 0.0001$; $\chi^2(2) = 305.092$.

stop codons as sense codons[1]. A similar departure from the standard genetic code was also documented in a dinoflagellate, *Amoebophrya*[2], and experimentally confirmed in the ciliates *Parduczia* and *Condylostoma*[3,4]. Although these findings challenge the basic principles of the code, the molecular mechanisms behind stop codon reassignments versus translation termination remain enigmatic.

Here we have identified a previously unknown molecular mechanism of the UGA reassignment to tryptophan in *B. nonstop* sp. nov. that has a universal character used by other unrelated eukaryotes. It combines a shortening of the $tRNA^{Trp}$ anticodon stem (AS) with a specific alteration of the eukaryotic release factor 1 (eRF1), leading to robustness, which was highly likely to be a prerequisite facilitating this (r)evolutionary reassignment.

Reassigned stop codons of *B. nonstop*

The trypanosomatid *B. nonstop* Votýpka et Lukeš, sp. nov., was isolated from the hindgut of a true bug, *Eysarcoris aeneus* (Hemiptera: Pentatomidae), captured in Podtrosecké rybníky, Czech Republic, and is considered here as a previously undescribed species on the basis of morphology and unique 18S rRNA sequence (Fig. 1a,b); for a detailed morphological description and phylogeny, see the Supplementary Taxonomic Summary and Extended Data Figs. 1 and 2a. We have sequenced the genome, transcriptome and proteome of *B. nonstop*, revealing that its genome is markedly G+C-low (G+C content is 35%) compared to an extensive collection of trypanosomatid genomes with an average G+C content of about 51% (Extended Data Fig. 2b). In agreement,

this comparison revealed that the genome of the common ancestor of the genus *Blastocrithidia* most probably experienced a severe A+T-mutational pressure generating the corresponding UGG-to-UGA, GAG-to-UAG, GAA-to-UAA, UGA-to-UAA and UAG-to-UAA codon substitutions (Extended Data Fig. 2b). Using mass spectrometry, we have confirmed the in silico-predicted meaning of the reassigned stop codons showing that UGA indeed encodes tryptophan, and UAG and UAA (UAR) specify glutamate (Fig. 1c,d and Extended Data Fig. 3a). In eukaryotes, the widespread nonsense-mediated decay pathway is responsible for the degradation of mRNAs with premature stop codons[20]. We propose that the loss of UPF1 and UPF2, the key components of this pathway, early in the evolution of Trypanosomatidae (Extended Data Fig. 2a), might have been one of the critical prerequisites of its stop codon reassignment.

As existing annotation programs could not handle ambiguous stop codons in *B. nonstop*, we developed an annotation algorithm relying on the alignments with the reference trypanosomatids and the position of spliced leader RNA *trans*-spliced on the 5′-end of all nuclear mRNAs (Methods and Extended Data Fig. 3b). We found that the frequency of the reassigned stop codons in protein-coding genes negatively correlates with the abundance of the corresponding proteins (Fig. 1e). For instance, mitochondrial ribosomal proteins expressed at low levels contain numerous in-frame UGA and UAR codons, whereas their heavily translated cytosolic homologues contain very few or none of these (Fig. 1f). Out of 7,259 predicted protein-coding genes in *B. nonstop*, only 228 lack any in-frame UGA and UAR codons (Extended Data Fig. 4a). Gene ontology analysis showed that they encode ribosomal proteins and other factors involved in cytosolic translation, histones, mitochondrial electron transport chain subunits, and other proteins expressed at high levels (Extended Data Fig. 4a).

Translation termination in *B. nonstop*

Previous bioinformatic analysis predicted that in *Blastocrithidia* sp., UGA, UAA and UAG are sense codons, with the last two also functioning as stop codons[1]. To identify genuine stop codons in the previously undescribed *B. nonstop*, we carried out BLASTp searches of the trypanosomatids' reference proteins (see Methods) and analysed the distribution of codons directly following the last codon of 3′-end-complete alignments (Fig. 1g). Most of these alignments terminated with UAA exceeding the next most abundant codon (AAA) more than 10 times. The abundance of the other two reassigned stop codons, UAG and UGA, was no different from that of other sense codons (Fig. 1g). In addition, we reanalysed the set of genes that was presented previously[1] as evidence that in *Blastocrithidia* sp. UAG may serve as a termination codon along with UAA. We found that in all of these cases, UAG was followed by UAA (Extended Data Fig. 4b), indicating that a handful of UAGs considered previously as termination codons most likely represent incorrect predictions. Moreover, we found that in *B. nonstop*, UAA was massively overrepresented after the 3′ ends of the protein-coding genes, whereas the two other stop codons were distributed uniformly (Fig. 1h). Markedly, for 1,569 genes for which we were able to define the 3′ untranslated regions (UTRs), this massive overrepresentation was observed specifically within the first 18 triplets downstream of the stop codon and only in the frame corresponding to the reading frame of a given gene (Fig. 1i and Extended Data Fig. 5). Together, these findings strongly indicate that UAA is used as the sole stop codon in *B. nonstop* and point to a peculiar termination mechanism.

Notably, and in contrast to the case for other trypanosomatids, our observations showed that the G+C distribution in the *B. nonstop* genome markedly differs inside the coding regions versus the intergenic regions (Extended Data Fig. 6a). The former is more G+C rich; however, immediately after the 3′ end of genes, the genome sequences become substantially more A+T rich. A sequence logo for the first 100 nucleotides after the stop codon shows that A is the most abundant nucleotide in the first ≈39 nt; after that T prevails (Extended Data Fig. 6b). These findings correspond with the use of UAA as the sole termination codon and, furthermore, indicate a possible involvement of poly(A)-binding protein (PABP), known to interact with A+U-rich sequences in other species[21–23], in the particularly challenging translation termination in *B. nonstop*.

UAR reassignment by cognate tRNAs^{Glu}

The nuclear genome of *B. nonstop* contains 70 tRNA genes, including tRNA$^{Glu}_{CUA}$ and tRNA$^{Glu}_{UUA}$ cognate to both UAR stop codons (Supplementary Data 1). Maximum-likelihood analysis showed that tRNA$^{Glu}_{CUA}$ is closely related to the canonical tRNA$^{Glu}_{CUC}$ and tRNA$^{Glu}_{UUA}$ is nested inside the tRNA$^{Glu}_{UUC}$ clade (Fig. 2a). Hence, phylogenetic analysis strongly supports the origin of these stop-codon-recognizing tRNAs from standard tRNAs^{Glu}. The expression and charging of both tRNA$^{Glu}_{CUA}$ and tRNA$^{Glu}_{UUA}$ in *B. nonstop* and their absence in the model species *T. brucei* were confirmed by northern blot analysis (Fig. 2b).

We found that a tRNA cognate to UGA encoding tryptophan is prominently missing from the *B. nonstop* genome, mirroring the situation in *C. magnum*[3]. Therefore, we reasoned that the canonical tRNA$^{Trp}_{CCA}$ might undergo CCA-to-UCA anticodon editing in the cytosol, as the same editing occurs in the mitochondrion of the closely related *Leishmania tarentolae*[24] and *T. brucei*[25]. We have recently identified a cytidine deaminase responsible for this mitochondrion-specific tRNA editing in *T. brucei*[26]. Although this enzyme is also present in *B. nonstop*, we demonstrate here that the corresponding editing event occurs exclusively in its mitochondrial but not cytosolic tRNA^{Trp} (Fig. 2c). In bacteria, an elevated UGA readthrough by tRNA^{Trp} is facilitated by the G24A mutation in the D-stem, known as the Hirsh suppressor[27], yet *B. nonstop* has a canonical G at the corresponding position 23 (Fig. 3a). Overall, sequencing of the cytosolic tRNAs from *B. nonstop* did not reveal any apparent post-transcriptional editing events that would allow increased UGA decoding as a sense codon[28].

UGA reassignment involves 4-bp tRNA^{Trp}

Secondary structure predictions using tRNAscan-SE[29] and ARAGORN[30] revealed that the AS of *B. nonstop* tRNA$^{Trp}_{CCA}$ is only 4 bp long, whereas the closely related *T. brucei* and other trypanosomatids possess a 5-bp-long AS of the canonical tRNA length (Fig. 3a). Moreover, we have discovered that ciliates with all three reassigned stop codons, such as *C. magnum*, also bear tRNA$^{Trp}_{CCA}$ with a shorter, 4-bp-long AS, in addition to the canonical 5-bp-long AS tRNA$^{Trp}_{CCA}$ (Fig. 3a). To the best of our knowledge, most other sequenced eukaryotes contain only the 5-bp-long stem variant of tRNA$^{Trp}_{CCA}$. Through a dedicated GtRNAdb search, we found a few tRNAs with a similarly shorter AS in organisms that are believed to have a standard set of stop codons. However, most of them seem to be tRNA-like pseudogenes with a questionable ability to recognize UGA. In any case, we posit that in the aforementioned protists tRNAs$^{Trp}_{CCA}$ exhibiting this unique feature may no longer tightly discriminate against the UGA stop codon and can reassign UGA (along with canonical UGG) to tryptophan.

As noted above, a double hydrogen bond formed between U26 and G42 in *T. brucei* tRNA$^{Trp}_{CCA}$ is missing in *B. nonstop* tRNA$^{Trp}_{CCA}$ owing to C26 and U42 replacements and, therefore, the AS of the latter tRNA is 1 bp shorter (Fig. 3a). We analysed the impact of this 4-bp-long AS on UGA decoding in cellulo by expressing an extra copy of the *B. nonstop* tRNA$^{Trp}_{CCA}$-encoding gene together with a dual-luciferase (DuaL) cassette with an in-frame UGA positioned between the firefly and *Renilla* luciferase genes in *T. brucei* (Extended Data Fig. 6c). The readthrough efficiency was determined by comparing the relative luciferase activities (firefly/*Renilla*) of the test constructs with respective controls, such as the CAA sense codon replacing UGA and cells lacking alien tRNA. The expression and aminoacylation status of each tRNA variant was verified by urea polyacrylamide gel electrophoresis (PAGE) and

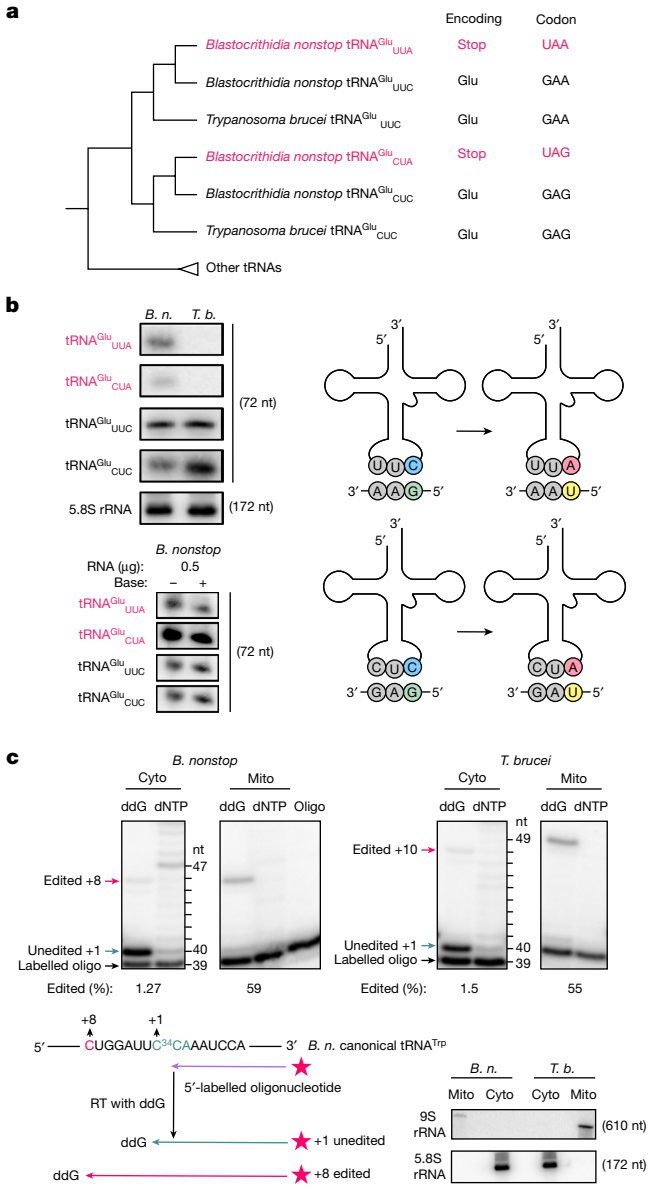


Fig. 2 | Phylogenetic analysis, structure, transcription and RNA editing of tRNA^{Glu} and tRNA^{Trp} of trypanosomatids. a, Cladogram summarizing the maximum-likelihood phylogenetic tree of tRNA^{Glu} of *B. nonstop* and related trypanosomatids showing that genes encoding tRNA$^{Glu}_{UUA}$ and tRNA$^{Glu}_{CUA}$ originated from standard tRNA$^{Glu}_{UUC}$ and tRNA$^{Glu}_{CUC}$, respectively. The full tree files are available in Supplementary Data 2. **b**, Left: northern blot analyses of tRNA$^{Glu}_{UUA}$ and tRNA$^{Glu}_{CUA}$, their expression and charging in *B. nonstop* (*B. n.*) and *T. brucei* (*T. b.*). The tRNA$^{Glu}_{UUC}$ and tRNA$^{Glu}_{CUC}$ genes, which are expressed in both organisms, served as controls. Right: schematic secondary structures for tRNAs^{Glu} recognizing all glutamate codons. Repeated three times with similar results. **c**, The 5′-^{32}P-radiolabelled oligonucleotide specific for tRNA $^{Trp}_{CCA}$ was used in a poisoned primer extension assay with cytosolic (cyto) and mitochondrial (mito) RNA fractions isolated from *B. nonstop* and *T. brucei*. Bands representing edited and unedited sites are indicated by the arrows; oligo, no template RNA was added to the primer extension reaction. The numbers below each autoradiogram represent relative amounts of the edited tRNA^{Trp}. Schematic of the poisoned primer extension assay is shown below. Here, the presence of ddG allows extension for unedited tRNA^{Trp} to +1 and edited tRNA^{Trp} to +8 (*B. nonstop*) or +10 (*T. brucei*), and +1 denotes the first position of the anticodon. The 9S rRNA and 5.8S rRNA (bottom, right) served as separation controls for mitochondrial and cytosolic RNA fractions, respectively; repeated five times with similar results. nt, nucleotides. For gel source data, see Supplementary Fig. 1.

urea–acid PAGE gels, respectively, followed by northern blot analysis using ^{32}P-radiolabelled tRNA-specific probes (Extended Data Fig. 7).

As predicted, the *T. brucei*-based DuaL assay revealed that expressing the *B. nonstop* tRNA$^{Trp}_{CCA}$ results in an approximately 8-fold higher readthrough over empty vector (EV), and a more than 2.3-fold increase over the genuine *T. brucei* tRNA$^{Trp}_{CCA}$ (Fig. 3b); the latter exhibited less than 3.5-fold increase over the EV control. The key evidence supporting our hypothesis came from the C26U and U42G substitutions in *B. nonstop* tRNA$^{Trp}_{CCA}$ restoring the 5-bp-long AS, thus mimicking tRNA$^{Trp}_{CCA}$ from other eukaryotes with the canonical genetic code. Readthrough levels produced by this restored tRNA mutant dropped down to those of the genuine *T. brucei* tRNA$^{Trp}_{CCA}$. Restoring the base pairing with a stronger pair at the fifth position, creating U42G substitution, resulted in an even weaker readthrough. The reciprocal proof of principle experiment, in which the fifth base pair of *T. brucei* tRNA$^{Trp}_{CCA}$ (U26C and G42U) was loosened to mimic the *B. nonstop* tRNA$^{Trp}_{CCA}$, showed twofold higher readthrough than its 5-bp-long wild-type tRNA (Fig. 3b). No such difference was observed for UAG and UAA (Extended Data Fig. 8a).

As a complementary approach to the luciferase activity assay, *Renilla* expression was directly assessed by western blotting of total lysates of *T. brucei* expressing *B. nonstop* tRNA$^{Trp}_{CCA}$ with either 4-bp- or 5-bp-long AS, along with the DuaL cassette with an in-frame UGA. Cells containing this cassette but lacking the alien tRNA and those expressing the DuaL cassette lacking in-frame UGA served as negative and positive controls, respectively (Fig. 3c). In clear agreement with our DuaL measurements, western blotting showed a strong stimulation of readthrough by the *B. nonstop* 4-bp-long AS tRNA$^{Trp}_{CCA}$ over its 5-bp-long variant. This is evidenced by the increased appearance of the prominent 98-kDa *Renilla*–firefly fusion protein at the expense of the 36-kDa signal corresponding to the *Renilla* protein alone (Fig. 3c).

To exclude a possible role of widespread tRNA modifications in the expanded decoding properties of tRNA$^{Trp}_{CCA}$, as reported for several other tRNAs[31,32], we carried out total nucleoside analysis of purified *B. nonstop* tRNA$^{Trp}_{CCA}$ by mass spectrometry. However, we have not detected any modifications that are known to facilitate UGA decoding in this and control species (Supplementary Information 2).

We next explored the attractive possibility that this phenomenon may be general and applicable to other organisms. Indeed, expressing the wild-type (5-bp-long stem) and mutated (U26C; 4-bp-long stem) tRNA$^{Trp}_{CCA}$ from *S. cerevisiae* in the DuaL system in *T. brucei* resulted in the same trend with a similar significance as in the case of *T. brucei* tRNA$^{Trp}_{CCA}$ and its shortened variant (Fig. 3b). Notably, analogous (4, high; 5, low) results were also obtained with two naturally coexisting tRNA$^{Trp}_{CCA}$ with a 4- versus 5-bp-long AS from *C. magnum*, which has also reassigned all three stop codons[3] (Fig. 3d).

Notably, mimicking this radical evolutionary twist in *S. cerevisiae* produced identical results. Loosening the fifth base pair of its tRNA$^{Trp}_{CCA}$ by the A42U or U26C, A42U substitutions in the *S. cerevisiae* DuaL system[33,34] significantly (≈3-fold) increased the UGA-tobacco mosaic virus (UGA-TMV; refers to UGA with the 3′ context of TMV as the TMV's genuine stop is UAG) readthrough over that of the wild-type 5-bp-long stem tRNA^{Trp}. The fact that the A42U substitution produced a similar result to that of U26C, A42U rules out a potential effect of the loss of the pseudouridine modification of U26C by the pseudouridylate synthase 1 (Pus1) enzyme[35]. Analogously, the wild-type *B. nonstop* tRNA$^{Trp}_{CCA}$ expressed in *S. cerevisiae* also facilitated robust fourfold readthrough stimulation over the wild-type *S. cerevisiae* tRNA^{Trp} in a clear dependence on the absence of the fifth base pair. This corresponds to 11-fold (for the 5-bp-long stem C26U, U42A restoration mutant) and 14-fold (for the 5-bp-long stem C26U, U42G restoration mutant) difference in favour of the 4-bp-long stem variant of *B. nonstop* tRNA$^{Trp}_{CCA}$ (Fig. 3e). Consistently and critically, the wild-type *T. brucei* tRNA$^{Trp}_{CCA}$ showed a weak readthrough unless its fifth base pair was broken through the U26C, G42U mutation, producing more than 3.5-fold increase in readthrough over the wild-type *S. cerevisiae* tRNA^{Trp} (Fig. 3e). This

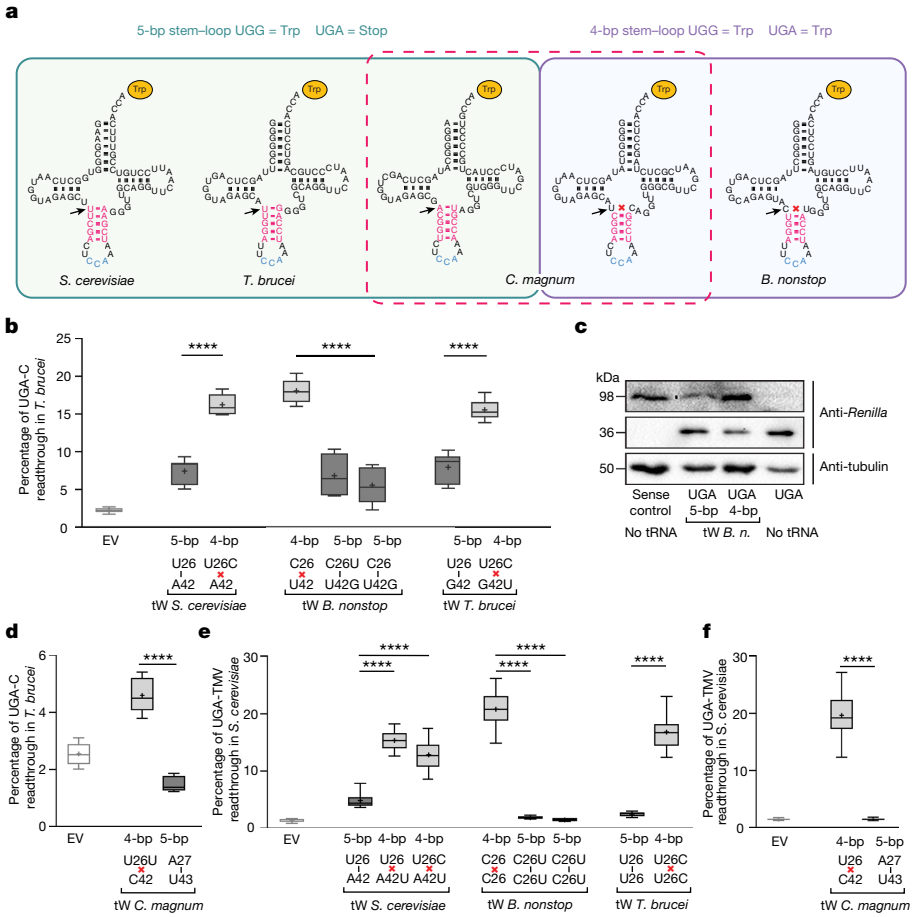


Fig. 3 | One base shorter AS of tRNA$^{Trp}_{CCA}$ is critical for UGA reassignment to tryptophan in *B. nonstop* and *C. magnum*. **a**, Predicted secondary structures of tRNA$^{Trp}_{CCA}$. ASs and anticodons are shown in magenta and blue, respectively. Arrows point to the fifth AS pair; crosses indicate lack of base pairing. **b**,**d**, The 4-bp AS of any tRNA$^{Trp}_{CCA}$ boosts UGA readthrough in *T. brucei* cells bearing dual-luciferase cassette with in-frame UGA. Cells were transformed with EV or tRNA$^{Trp}_{CCA}$ (tW) with 4-bp or 5-bp AS versions from *S. cerevisiae, B. nonstop, T. brucei* (**b**) and *C. magnum* (**d**), and processed for UGA readthrough measurements as described in Methods. Readthrough values were normalized to those of the control cell line (containing dual-luciferase cassette without in-frame UGA). Each box represents $n = 9$ values (3 individual experiments each including 3 biological replicates); whiskers range from minimal to maximal values; plus symbol indicates the mean value; vertical black line, 26·42 base pairing; red cross, no pairing. Statistical significance was determined by the unpaired, two-tailed Welch's *t*-test; ****$P < 0.0001$. **c**, Western blot analysis of *Renilla* expression using anti-*Renilla* antibodies recognizing both the *Renilla*–firefly fusion protein (98 kDa) and the *Renilla* protein (36 kDa); repeated twice. **e**,**f**, The 4-bp AS of any tRNA$^{Trp}_{CCA}$ boosts UGA-TMV readthrough in yeast. H541 was transformed with a corresponding DuaL readthrough reporter together with an EV or a given high-copy tRNA^{Trp} variant. Transformants were processed for UGA readthrough measurements and analysed as in **b**; ****$P < 0.0001$; in **e** each box represents $n = 9$ values (3 individual experiments each including 3 biological replicates); in **f** each box represents $n = 18$ values (3 individual experiments each including 6 biological replicates). For gel source data, see Supplementary Fig. 1.

corresponds to sevenfold difference between the 5- versus 4-bp-long stem variants of *T. brucei* tRNA$^{Trp}_{CCA}$. Such a difference was observed neither on UAG and UAA stop codons nor under amino acid starvation stress induced following treatment with 3-aminotriazol (Extended Data Fig. 8b–d). To rule out that the observed differences are due to different levels of expressed tRNA variants, we confirmed their comparable abundance and aminoacylation by urea PAGE and urea–acid PAGE, respectively, followed by northern blot analysis using tRNA-specific probes (Extended Data Fig. 9). We also obtained data quantitatively and qualitatively very similar to our measurements with *B. nonstop* tRNA$^{Trp}_{CCA}$ using two naturally occurring tRNA$^{Trp}_{CCA}$ from *C. magnum* (Fig. 3f), although here—in contrast to the case for *T. brucei*—charging of the *C. magnum* 5-bp variant in *S. cerevisiae* is not so obvious.

In *S. cerevisiae*, the outcome of these measurements was not affected by the nature of the stop codon tetranucleotide represented by the stop codon and the base immediately following it, which plays a critical role in termination[36,37] (Extended Data Fig. 8e). For all tested tRNAs, we observed increases of several fold corresponding to the specificity of *S. cerevisiae* wild-type tRNA^{Trp} for individual tetranucleotides, as shown previously[34]. Notably, this result also rules out a potential effect of the artificial AS shortening on the Trm7-mediated 2′-*O*-ribose methylation of tRNA$^{Trp}_{CCA}$ at positions C32 and C34 (the latter being the wobble position of the anticodon[35]), because in a *trm7Δ* background, readthrough by tRNA$^{Trp}_{CCA}$ is specifically increased only at the UGA-C and UGA-A tetranucleotides[34]. These findings suggest that we detected a specific effect of the length of the AS and not of the anticodon per se.

Genetic manipulations shortening the AS from 5 bp to 4 bp of two other tRNAs near-cognate to UGA, namely tRNA$^{Cys}_{GCA}$ and tRNA$^{Arg}_{UCU}$, did not alter the termination efficiency in *S. cerevisiae* in any significant way (Extended Data Fig. 8f). Stable overexpression of the wild type variants of these two tRNAs was documented previously[33,34].

UGA reassignment involves mutated eRF1 variant

eRF1 from *Blastocrithidia* sp. was reported to bear a specific Ser74Gly substitution (in the *B. nonstop* numbering; Ser67 in *S. cerevisiae*) in

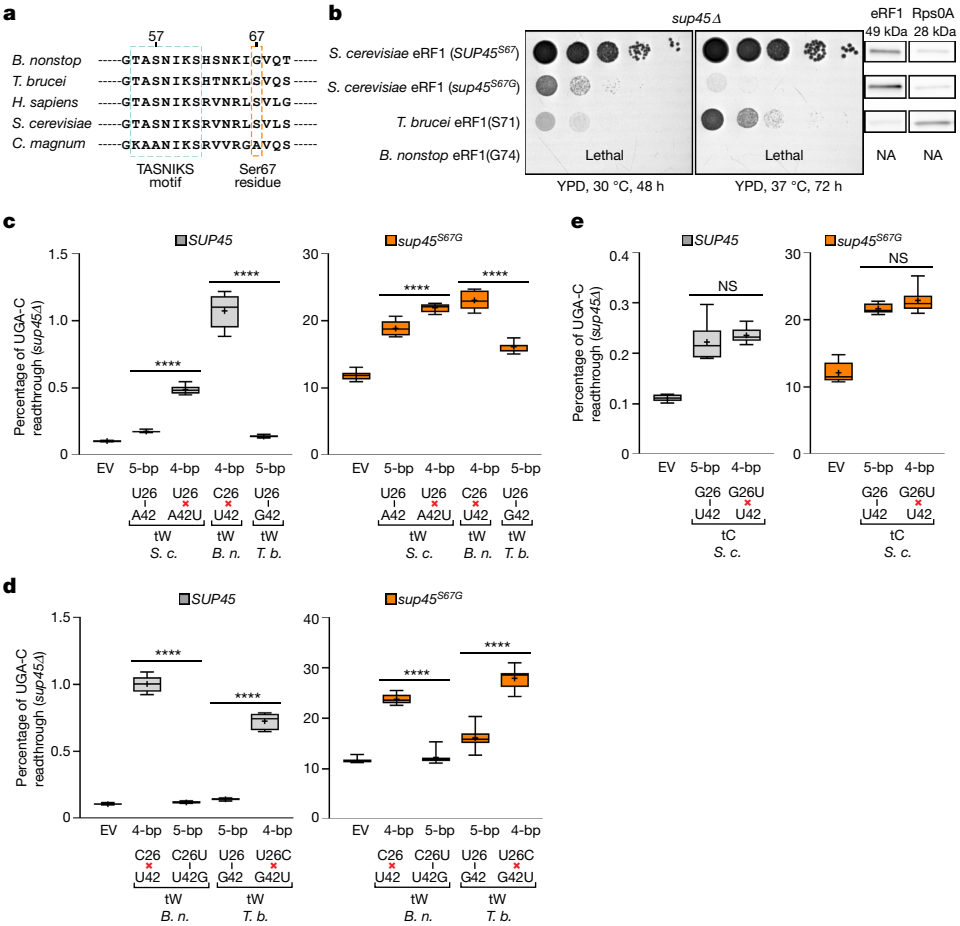


Fig. 4 | Specific eRF1 alteration potentiates reassignment of UGA as tryptophan. a, Sequence alignment of eRF1 (residues 54–70 *S. cerevisiae* numbering) from indicated organisms. The conserved TASNIKS motif promoting stop codon recognition (residues 55–61) and residues occurring at the critical position 67 are indicated. *H. sapiens*, *Homo sapiens*. **b**, The yeast eRF1 mutant *SUP45*^{S67G} and *T. brucei eRF1*^{S71} expressed in yeast impart severe slow growth phenotypes. Yeast strains ZH193 (*sup45Δ* + *SUP45*), ZH238 (*sup45Δ* + *SUP45*^{S67G}) and ZH235 (*sup45Δ* + eRF1 of *T. brucei*) were spotted in five serial tenfold dilutions on YPD medium and incubated for 48 h at 30 °C and for 72 h at 37 °C; western blot analyses, carried out with samples run on the same gel in two biological replicates, are shown on the right. **c**, Yeast eRF1(S67G) robustly potentiates UGA reassignment by the *B. nonstop* 4-bp-long AS tRNA^{Trp}. Yeast strains bearing wild-type eRF1 (*sup45Δ* + *SUP45*) or eRF1(S67G) (*sup45Δ* + *SUP45*^{S67G}) were transformed with a corresponding DuaL readthrough reporter together with an EV or a given high-copy tRNA^{Trp} variant. Transformants were treated and analysed, and are plotted and depicted, as in Fig. 3; ****$P < 0.0001$; each box represents $n = 12$ values (3 individual experiments each including 4 biological replicates). *S. c.*, *S. cerevisiae*. **d**, Readthrough levels of 4-bp- versus 5-bp-long AS variants of *B. nonstop* versus *T. brucei* tRNA^{Trp} measured in the background of yeast eRF1(S67G); same as in **c**, including the *n* value. **e**, Readthrough levels of 4-bp- versus 5-bp-long AS variants of yeast tRNA^{Cys} (also near-cognate to UGA) measured in the background of yeast eRF1(S67G); same as in **c**, including the *n* value, except that tRNA^{Cys} (tC) variants were tested; ****$P < 0.0001$. For gel source data, see Supplementary Fig. 1.

an otherwise highly conserved position[1] (Fig. 4a and Extended Data Fig. 10). When a similar Ser67Ala substitution was introduced into yeast or human eRF1 homologues and expressed in *S. cerevisiae*, a significantly increased readthrough over UGA but not at all over UAR was detected[38]. This clearly demonstrates that Ser67Ala specifically restricts UGA decoding in vivo, as observed in vitro for eRF1 from the ciliate *Euplotes aediculatus*[39].

To test for genetic interaction between this *B. nonstop*-specific Gly residue of eRF1 and the stem length variants of tRNA^{Trp}, we substituted the original Ser67 with Gly in the *S. cerevisiae* eRF1 (encoded by *SUP45*). We first revealed that it severely reduced cellular growth (Fig. 4b) and confirmed restriction of UGA decoding in vivo (Extended Data Fig. 8g). Then, we used the *S. cerevisiae* reporter system described above. First, in cells deleted for *sup45* expressing the wild-type eRF1 from a plasmid along with the UGA-C reporter, we recapitulated the 4- versus 5-bp-long AS data with both wild-type and mutated *S. cerevisiae* tRNA^{Trp}, as well as with naturally occurring tRNA^{Trp} of *T. brucei* and *B. nonstop* (Fig. 4c; left panel). Second, the Ser67Gly mutant markedly (by >100-fold) increased readthrough over the wild type expressing an EV (Fig. 4c; compare EV between both panels). Such a robust increase unambiguously explains the severity of the observed growth defect (Fig. 4b). Third, in spite of this marked increase, the readthrough values over UGA-C further increased with all tRNA^{Trp} variants tested in the Ser67Gly cells, preserving the statistically significant 4- versus 5-bp-long AS trend observed in wild-type cells (Fig. 4c). The fold differences between 4- and 5-bp-long stem variants were smaller in the mutant, reflecting the robustness of the Ser67Gly-mediated readthrough. However, it must be noted that the relatively smaller 1.5-fold difference between *B. nonstop* and *T. brucei* tRNA^{Trp} variants in Ser67Gly cells actually means 220-fold versus 150-fold more efficient readthrough over the EV wild-type background, respectively, which, in absolute terms, represents yet another robust increase.

Direct comparison of 4- and 5-bp-long stem variants of tRNA^{Trp} of *T. brucei* and *B. nonstop* in *S. cerevisiae* expressing wild-type or mutant eRF1 further illustrated the power of this particular combination. We observed ≈2-fold difference between 4- versus 5-bp-long stem variants of *B. nonstop* tRNA^{Trp} in Ser67Gly cells (that is, 220-fold versus 110-fold over EV, respectively; Fig. 4d). This functionally independent

multiplicative phenotype suggests that the natural Ser74Gly substitution in *B. nonstop* eRF1 and the shortened $tRNA^{Trp}$ AS work in concert to maximize UGA reassignment to tryptophan. As a specificity control, the 4- versus 5-bp-long stem variants $tRNA^{Cys}_{GCA}$ (also near-cognate to UGA) showed no differences in both strains (Fig. 4e).

Last, our attempt to exploit the *S. cerevisiae* system to express and directly test the *T. brucei* and *B. nonstop* eRF1 variants revealed that, whereas expressing *T. brucei* eRF1 as the sole source of eRF1 causes severe slow growth, expressing *B. nonstop* eRF1 is lethal (Fig. 4b). We propose that the reason for this is the robust readthrough potential of *B. nonstop* eRF1 that exceeds the tolerance threshold to sustain life in an organism relying on the standard genetic code.

Discussion

Reassignment of all three stop codons evolved in evolutionary distant eukaryotes several times independently[2–4,7], yet the reasons for, and evolutionary mechanisms behind, such a massive departure from the canonical genetic code remained largely unknown. Here we isolated a previously undescribed trypanosomatid *B. nonstop* and dissected and characterized the molecular basis of all necessary adaptations allowing this parasitic flagellate to survive under such non-canonical conditions.

Multiple unique traits of the *B. nonstop* genome are consistent with the assumption that the ancestor of *Blastocrithidia* spp. experienced a severe A+T-mutational pressure, making G+C-to-A+T substitutions frequent. The most deleterious consequence of this shift is the appearance of nonsense mutations, causing premature translation termination, inevitably compromising the function of produced truncated proteins. We revealed here that to counteract this undesirable phenomenon, the ancestor of *Blastocrithidia* spp. has evolved independent mechanisms to: prevent the occurrence of in-frame stop codons in and/or eliminate them from genes expressed at high levels; set UAA as the sole stop codon; and adapt corresponding tRNAs and eRF1 to reassign in-frame stop codons as sense codons (Supplementary Video 1). These are individually discussed in the following paragraphs.

Negative correlation of decreasing representation of in-frame stop codons with increasing protein abundance strongly indicates a hidden burden associated with their presence in mRNAs translated at high rates. Indeed, proteins involved in essential processes that, if under-expressed owing to the abortive, stalled or imprecise translation, may cause serious cell damage or death, are entirely devoid of the in-frame stop codons. Whether they were prevented from appearing (more likely) or promptly eliminated remains to be elucidated (see extended discussion in Supplementary Information for more details).

Our bioinformatic analysis revealed that, in disagreement with the earlier report on *Blastocrithidia* sp.[1], UAA is used as the sole stop codon in *B. nonstop*, which distinguishes them from ciliates[13] (see extended discussion on the termination mechanism in Supplementary Information). Notably, UAA is the only significantly overrepresented codon in the sequence encompassing the first 18 triplets after the genuine stop codon, but solely in the 'coding' frame of a given gene (Fig. 1i and Extended Data Fig. 5); it becomes equally abundant in all three frames afterwards. It is reasonable to assume that following the emergence of $tRNA^{Glu}$ fully cognate to UAA, the 'run-out' coding frames were subjected to positive selection, allowing the formation of a battery of in-frame UAAs immediately following the genuine UAA stop codon to counteract the expected termination interference by this non-canonical $tRNA^{Glu}$.

As the reassignment of UAR stop codons by two new fully cognate tRNAs documented here in *B. nonstop* has been previously described in ciliates[3,4], the most striking finding is the shortening of the canonical 5-bp-long AS of $tRNA^{Trp}_{CCA}$ to a 4-bp-long variant, as an adaptation of the trypanosomatid to the in-frame stop codons. We demonstrated that this relatively minor alteration allowed *B. nonstop* to gain a robustly increased UGA readthrough potential. Not only is this phenomenon experimentally transferrable to *S. cerevisiae* and the human parasite *T. brucei*, but we also showed that the evolutionary unrelated ciliate *C. magnum* has independently adopted the same strategy. Mechanistically, our findings suggest that the decoding A site of the ribosome can monitor the geometry of not only the codon–anticodon heteroduplex but also of the tRNA AS during the codon sampling and accommodation phases of elongation (see extended discussion in Supplementary Information for more details). The rationale behind the inability of these organisms to generate $tRNA^{Trp}_{UCA}$ fully cognate to UGA, as in the case of glutamate tRNAs and UAR stop codons, may rest in the fact that tryptophanyl-tRNA synthetase is anticodon-sensitive and could not charge the $tRNA^{Trp}_{UCA}$, even if it did evolve[40,41], unlike, for instance, glutamyl-tRNA synthetase and glutamate-tRNA ligase[42].

On the basis of our eRF1 mutagenesis, it seems plausible that shortening the $tRNA^{Trp}$ AS may not have been sufficient to achieve the level of UGA reassignment required for the survival of *B. nonstop*. Hence, evolutionary pressure eventually resulted in selection of an organism with two independent mutations, one in the $tRNA^{Trp}$ and the other in its eRF1-encoding gene, the latter compromising the ability of eRF1 to recognize UGA. As a result, even the canonical 5-bp-long AS variant of $tRNA^{Trp}_{CCA}$ can robustly overpower eRF1 in its competition for UGA, producing a striking ≈100-fold increase in the UGA readthrough. In the case of the 4-bp-long AS variant, the UGA readthrough rockets to >200-fold. The fact that evolutionary distant ciliates with reassigned UGA acquired a similar eRF1 substitution (Ser69Ala in *C. magnum* numbering)[39] provides strong support for coevolution of these two critical adaptations, the combination of which allows sufficiently robust UGA reassignment (see Supplementary Video 1 and Supplementary Discussion for more details). Notably, experiments with mutated yeast eRF1 (Extended Data Fig. 8g)[38], as well as in vitro studies with mutant eRF1 from ciliates[39], revealed that, in contrast to the case for UGA, the recognition of UAR codons by eRF1 mutated at this particular position remains unaffected. This strongly suggests that in *B. nonstop* the recognition of UAR codons remains unaffected as well, which may explain the evolutionary pressure for their complete reassignment to glutamate by fully cognate tRNAs.

Reflecting on the apparent initial disadvantage(s) of alternative genetic codes opens an important question regarding their benefits. Considering that the affected organisms are successful parasites with a cosmopolitan distribution, this putative disadvantage must be very subtle and/or outweighed by some advantages. One such unexpected benefit was recently demonstrated in *Escherichia coli*, in which codon reassignment led to its resistance to a cocktail of otherwise pathogenic bacteriophages[43]. As trypanosomatids frequently host various deleterious viruses[44], a total stop codon reassignment may represent an elegant way to mitigate their virulence.

As *B. nonstop* utilizes only a single genuine stop codon, it liberates the other two reassigned stop codons for ordinary coding purposes. This way it achieved one of the aims of synthetic biology, namely to reconfigure the genetic code to free codons for the incorporation of the non-canonical amino acids[45]. As the features described in this obscure flagellate are transferrable to other eukaryotes, our research opens a previously unexplored route in genetic code manipulations. For example, one of the biotechnological implications of our findings could be to optimize the $SUP45^{S67G}$ *S. cerevisiae* strain overexpressing the 4-bp-long AS variant of $tRNA^{Trp}_{CCA}$ to allow efficient stop codon reassignment of interest. Moreover, this discovery may provide a new way for the design of suppressor tRNAs for the treatment of inherited diseases, as recently enabled through a recombinant adeno-associated virus gene delivery vehicle in mice[46].

Taking these findings together, we conclude that a subtle reconfiguration of $tRNA^{Trp}$ in a parasitic protist, transmissible to $tRNAs^{Trp}$ across the eukaryotic supergroups, coupled with a functional alteration of eRF1 and duplication and neofunctionalization of $tRNA^{Glu}$ allowed a wholesale stop codon reassignment, the potential of which can now be explored in other genetically tractable organisms.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05584-2.

1. Záhonová, K., Kostygov, A. Y., Ševčíková, T., Yurchenko, V. & Eliáš, M. An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. *Curr. Biol.* **26**, 2364–2369 (2016).
2. Bachvaroff, T. R. A precedented nuclear genetic code with all three termination codons reassigned as sense codons in the syndinean *Amoebophrya* sp. ex *Karlodinium veneficum*. *PLoS ONE* **14**, e0212912 (2019).
3. Swart, E. C., Serra, V., Petroni, G. & Nowacki, M. Genetic codes with no dedicated stop codon: context-dependent translation termination. *Cell* **166**, 691–702 (2016).
4. Heaphy, S. M., Mariotti, M., Gladyshev, V. N., Atkins, J. F. & Baranov, P. V. Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in *Condylostoma magnum*. *Mol. Biol. Evol.* **33**, 2885–2889 (2016).
5. Sella, G. & Ardell, D. H. The coevolution of genes and genetic codes: Crick's frozen accident revisited. *J. Mol. Evol.* **63**, 297–313 (2006).
6. Koonin, E. V. & Novozhilov, A. S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **61**, 99–111 (2009).
7. Lobanov, A. V. et al. Position-dependent termination and widespread obligatory frameshifting in *Euplotes* translation. *Nat. Struct. Mol. Biol.* **24**, 61–68 (2017).
8. Shulgina, Y. & Eddy, S. R. A computational screen for alternative genetic codes in over 250,000 genomes. *eLife* **10**, e71402 (2021).
9. Keeling, P. J. Evolution of the genetic code. *Curr. Biol.* **26**, R851–R853 (2016).
10. Baranov, P. V., Atkins, J. F. & Yordanova, M. M. Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. *Nat. Rev. Genet.* **16**, 517–529 (2015).
11. Keeling, P. J. & Leander, B. S. Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. *J. Mol. Biol.* **326**, 1337–1349 (2003).
12. Karpov, S. A. et al. Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi. *Protist* **164**, 195–205 (2013).
13. Lozupone, C. A., Knight, R. D. & Landweber, L. F. The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.* **11**, 65–74 (2001).
14. Sanchez-Silva, R., Villalobo, E., Morin, L. & Torres, A. A new noncanonical nuclear genetic code: translation of UAA into glutamate. *Curr. Biol.* **13**, 442–447 (2003).
15. Osawa, S. & Jukes, T. H. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **28**, 271–278 (1989).
16. Schultz, D. W. & Yarus, M. Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.* **235**, 1377–1380 (1994).
17. Sengupta, S. & Higgs, P. G. A unified model of codon reassignment in alternative genetic codes. *Genetics* **170**, 831–840 (2005).
18. Lukeš, J. et al. Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. *Trends Parasitol.* **34**, 466–480 (2018).
19. Maslov, D. A. et al. Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology* **146**, 1–27 (2019).
20. He, F. & Jacobson, A. Nonsense-mediated mRNA decay: degradation of defective transcripts is only part of the story. *Annu. Rev. Genet.* **49**, 339–366 (2015).
21. Baejen, C. et al. Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. *Mol. Cell* **55**, 745–757 (2014).
22. Kini, H. K., Silverman, I. M., Ji, X., Gregory, B. D. & Liebhaber, S. A. Cytoplasmic poly(A) binding protein-1 binds to genomically encoded sequences within mammalian mRNAs. *RNA* **22**, 61–74 (2016).
23. Sladic, R. T., Lagnado, C. A., Bagley, C. J. & Goodall, G. J. Human PABP binds AU-rich RNA via RNA-binding domains 3 and 4. *Eur. J. Biochem.* **271**, 450–457 (2004).
24. Alfonzo, J. D., Blanc, V., Estevez, A. M., Rubio, M. A. & Simpson, L. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J.* **18**, 7056–7062 (1999).
25. Wohlgamuth-Benedum, J. M. et al. Thiolation controls cytoplasmic tRNA stability and acts as a negative determinant for tRNA editing in mitochondria. *J. Biol. Chem.* **284**, 23947–23953 (2009).
26. Paris, Z. et al. A mitochondrial cytidine deaminase is responsible for C to U editing of tRNA(Trp) to decode the UGA codon in *Trypanosoma brucei*. *RNA Biol.* **18**, 278–286 (2021).
27. Hirsh, D. Tryptophan transfer RNA as the UGA suppressor. *J. Mol. Biol.* **58**, 439–458 (1971).
28. Nenarokova, A. & Paris, Z. tRNAseq analysis of *Blastocrithidia nonstop*. *figshare* https://doi.org/10.6084/m9.figshare.17934200.v2 (2022).
29. Chan, P. P. & Lowe, T. M. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1–14 (2019).
30. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
31. Van Haute, L., Powell, C. A. & Minczuk, M. Dealing with an unconventional genetic code in mitochondria: the biogenesis and pathogenic defects of the 5-formylcytosine modification in mitochondrial tRNA(Met). *Biomolecules* **7**, 24 (2017).
32. Agris, P. F. et al. Celebrating wobble decoding: half a century and still much is new. *RNA Biol* **15**, 537–553 (2018).
33. Beznosková, P., Gunisová, S. & Valášek, L. S. Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. *RNA* **22**, 456–466 (2016).
34. Beznosková, P., Pavlíková, Z., Zeman, J., Echeverria Aitken, C. & Valášek, L. S. Yeast applied readthrough inducing system (YARIS): an *in vivo* assay for the comprehensive study of translational readthrough. *Nucleic Acids Res.* **47**, 6339–6350 (2019).
35. Pineyro, D., Torres, A. G. & de Pouplana, L. R. In *Fungal RNA Biology* (eds Sesma, A. & von der Haar, T.) 233–267 (Springer, 2014).
36. Matheisl, S., Berninghausen, O., Becker, T. & Beckmann, R. Structure of a human translation termination complex. *Nucleic Acids Res.* **43**, 8615–8626 (2015).
37. Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. *Nature* **524**, 493–496 (2015).
38. Blanchet, S. et al. New insights into stop codon recognition by eRF1. *Nucleic Acids Res.* **43**, 3298–3308 (2015).
39. Eliseev, B., Kryuchkova, P., Alkalaeva, E. & Frolova, L. A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. *Nucleic Acids Res.* **39**, 599–608 (2011).
40. Xue, H., Shen, W., Giege, R. & Wong, J. T. Identity elements of tRNA(Trp). Identification and evolutionary conservation. *J. Biol. Chem.* **268**, 9316–9322 (1993).
41. Ulmasov, B., Topin, A., Chen, Z., He, S. H. & Folk, W. R. Identity elements and aminoacylation of plant tRNATrp. *Nucleic Acids Res.* **26**, 5139–5141 (1998).
42. Sekine, S. et al. Major identity determinants in the "augmented D helix" of tRNA(Glu) from *Escherichia coli*. *J. Mol. Biol.* **256**, 685–700 (1996).
43. Robertson, W. E. et al. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
44. Grybchuk, D. et al. Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite *Leishmania*. *Proc. Natl Acad. Sci. USA* **115**, E506–E515 (2018).
45. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
46. Wang, J. et al. AAV-delivered suppressor tRNA overcomes a nonsense mutation in mice. *Nature* **604**, 343–348 (2022).

Methods

T. brucei and *S. cerevisiae* strains and plasmids

The lists and descriptions of all strains, plasmids, primers and GeneArt Strings DNA Fragments (Invitrogen) used in this study can be found in Supplementary Tables 1–4.

Northern blot and primer extension analysis of *B. nonstop* and *T. brucei*

Total RNA from *B. nonstop* and *T. brucei* was isolated with Tri-Reagent (Sigma) using the manufacturer's protocol. Ten micrograms of total RNA was separated on denaturing 8% PAGE gel with 8 M urea and electroblotted to Zeta-probe (Bio-Rad) membranes, which were subsequently probed with oligonucleotides specific for each RNA labelled with ^{32}P at their 5′ end (Supplementary Table 3). Hybridization procedures were carried out according to the manufacturer's instructions (Bio-Rad). Images were taken with a Storm PhosphoImager (Molecular Dynamics). Primer extension analysis was carried out according to the method described previously[24].

Acid PAGE and northern blot analysis of *T. brucei*

A total of 4×10^8 cells of *T. brucei* bearing alien tRNA variants were collected by centrifugation at 1,300*g* for 10 min at 4 °C. The resulting cell pellet was resuspended in 0.3 M of sodium acetate pH 4.5 and 10 mM EDTA, and RNA was extracted by adding equal volumes of phenol and chloroform, pH 4.7, followed by ethanol precipitation. The deacylated control RNA was obtained by incubation in 100 mM Tris-HCl, 100 mM NaCl, pH 9.5. The PAGE electrophoresis was carried out in 0.3 M sodium acetate, pH 5.0, at 4 °C at 50 V for 24 h, followed by northern blotting.

Acid PAGE and northern blot analysis of *S. cerevisiae*

To examine the level of tRNA expression, RNA isolated by the Quick RNA miniprep assay was separated using Criterion Precast Gels (Bio-Rad) and transferred onto 0.45-µm nylon transfer membrane (Bio-Rad). Northern blot analysis was carried out according to the DIG Northern Starter Kit protocol (Roche). Custom-made DIG-labelled oligonucleotides (Supplementary Table 3) were used as probes.

To examine the level of tRNA aminoacylation, total RNA was prepared under acidic conditions and stored in 10 mM sodium acetate buffer, pH 4.7, to preserve charged tRNAs. Aliquots of about 50 µg were subjected to deacylation as described elsewhere[47]. Charged and uncharged tRNAs were resolved in acid–urea 12% PAGE gels, blotted and hybridized as described above.

Stop codon readthrough assays

Most stop codon readthrough assays were carried out using a standard bicistronic reporter construct bearing a *Renilla* luciferase gene followed by an in-frame firefly luciferase gene, developed previously[48]. The two genes are separated by either a tetranucleotide termination signal (UGA-C) or, for control purposes, the CAA sense codon, followed by cytosine. In indicated cases, the termination signal and/or the following nucleotide context was modified. Note that mRNA levels of the reporters bearing the stop signal between *Renilla* and firefly genes do not differ from the CAA sense control[34]. This system avoids possible artefacts connected to the changes in the efficiency of translation initiation associated with the nonsense-mediated decay pathway[49], because both *Renilla* and firefly enzymes initiate translation from the same AUG codon[50]. All experiments and data analyses were carried out according to the microtitre-plate-based DuaL luciferase assay (Promega). Readthrough measurements were conducted from at least three biological replicates ($n \geq 3$), and each experiment was repeated at least three times. The readthrough values are represented by the box-and-whisker plot with the mean value marked (plus symbol) and whiskers going from minimal to maximal value. The box always extends from the 25th to 75th percentiles. The line in the middle of the box is plotted at the median. Statistical significance was determined by the unpaired, two-tailed Welch's *t*-test. The readthrough calculations with raw data for firefly and *Renilla* measurements are provided in the Source Data for Figs. 3 and 4 and Extended Data Fig. 8.

Total nucleoside analysis of single purified $tRNA^{Trp}$ (liquid chromatography–mass spectrometry)

The single purified $tRNA^{Trp}_{CCA}$ was isolated from the cytosolic RNA of *B. nonstop* and *T. brucei* using a specific 5′ biotinylated oligonucleotide probe (Supplementary Table 3) and followed by digestion of the purified tRNA fraction as described previously[51]. Separation was accomplished by reversed-phase chromatography using an Acquity UPLC HSS T3, 1.8 µm, 1 mm × 100 mm column (Waters) on a Vanquish Flex Quaternary UHPLC system (Thermo Fisher Scientific). The mobile phase A consisted of 5.3 mM ammonium acetate in liquid chromatography–mass spectrometry (MS)-grade water (Alfa Aesar), pH 5.3. The mobile phase B consisted of a 60:40 mixture of 5.3 mM ammonium acetate and acetonitrile (Honeywell Burdick & Jackson) with a gradient of 0% B (from 0 to 1.8 min), 2% B at 3 to 3.5 min, 3% B at 4.1 min, 5% B at 7 min, 25% B at 9 min, 35% B at 15 min, 99% B at 15.5 min (hold for 4.5 min), and 99% B at 20 min, and then returning to 0% B at 25.5 min, at a flow rate of 100 µl min^{-1}. The column temperature was set to 40 °C. Accurate mass values and physio-chemical properties were determined using Marvin v. 17.3.13.0 (https://chemaxon.com).

High-resolution accurate mass analysis was carried out on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) interfaced with an H-ESI electrospray source in positive polarity mode. Full scan data were acquired at a resolution of 120,000, mass range 220–900 *m*/*z*, AGC 7.5×10^4 and IT 100 ms. Data-dependent top speed MS/MS spectra (1-s cycle, CID 40%) were acquired in the ion trap with rapid scan rate, AGC 1.0×10^5 and an injection time of 300 ms, with scan range mode set to auto *m*/*z* normal. The other instrumental conditions were as follows: quadrupole isolation of 1 *m*/*z*; RF 35%; sheath gas, auxiliary gas and sweep gas of 30, 10 and 0 arbitrary units, respectively; ion transfer tube temperature of 289 °C; vaporizer temperature of 92 °C; and spray voltage of 3,500 V. Data were analysed using Xcalibur 4.0, Compound Discoverer 3.0 and mzVault 2.1 (Thermo Fisher Scientific).

Whole-cell extract preparation and western blot analysis

S. cerevisiae strains bearing wild-type or mutant eRF1 were cultured in YPD medium and collected at an optical density of about 1.0 at 600 nm by centrifugation at 1,400*g* for 5 min at 4 °C. The pellet was resuspended in 10 ml of ice-cold water and collected by centrifugation as above. The cell pellet was weighed and resuspended in 1:1 ratio (g:ml) in GF buffer, pH 7.5 (20 mM Tris-HCl, 50 mM KCl, 10 mM $MgCl_2$, 5 mM NaF, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1 µg ml^{-1} aprotinin, 1 µg ml^{-1} leupeptin, 1 µg ml^{-1} pepstatin and EDTA-free Complete Protease Inhibitor Mix). The whole-cell extracts were prepared by breaking the cells in 1.5-ml tubes using glass beads (about 2/3 volume) by the Bead Beater Homogenizer. After centrifugation at 15,800*g* for 5 min at 4 °C, the supernatant was transferred into a pre-cooled 1.5-ml tube and cleared by two consecutive rounds of centrifugation at 15.871*g* at 4 °C for 2 and 10 min. The total protein content was measured by the Bradford assay. A 10-µg aliquot of protein samples was resolved by SDS–PAGE, transferred onto a nitrocellulose membrane and incubated with anti-eRF1 (ref. [52]; at 1:500) and anti-Rps0A antibodies[53] (at 1:500). The western blot was developed using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific).

Various transfected strains of *T. brucei*, collected 24 h following the induction of expression from the dual-luciferase cassette, were collected at a density of 5×10^6 cells ml^{-1}, lysed and loaded on a 15% SDS–PAGE gel. PVDF membranes with transferred proteins were incubated with anti-*Renilla* (Merck) and anti-tubulin mouse antibodies (Thermo Fischer Scientific) at 1:500 and 1:7,000 dilutions, respectively. The membranes were subsequently incubated with anti-mouse

horseradish-peroxidase-coupled secondary antibody (Sigma-Aldrich) at 1:2,000 and visualized using Clarity Western ECL Substrate (Bio-Rad).

Bioinformatics

A detailed description of the bioinformatic methods used in this study is provided in the Supplementary Information.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated during this study are included in this published article (and its Supplementary Information files). The MS data have been deposited to the ProteomeXchange Consortium through the PRIDE[54] partner repository with the dataset identifier PXD033324. The high-throughput sequencing datasets were deposited in the National Center for Biotechnology Information under the number PRJNA790628 and in Figshare[55]. Additional data and analyses are available in Figshare[56]. Source data are provided with this paper.

Code availability

Custom computer code that was written for this project is publicly available on Zenodo[57].

47. Janssen, B. D., Diner, E. J. & Hayes, C. S. Analysis of aminoacyl- and peptidyl-tRNAs by gel electrophoresis. *Methods Mol. Biol.* **905**, 291–309 (2012).
48. Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479–486 (1998).
49. Muhlrad, D. & Parker, R. Recognition of yeast mRNAs as "nonsense containing" leads to both inhibition of mRNA translation and mRNA degradation: implications for the control of mRNA decapping. *Mol. Biol. Cell* **10**, 3971–3978 (1999).
50. Loughran, G., Howard, M. T., Firth, A. E. & Atkins, J. F. Avoidance of reporter assay distortions from fused dual reporters. *RNA* **23**, 1285–1289 (2017).
51. Ross, R., Cao, X., Yu, N. & Limbach, P. A. Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* **107**, 73–78 (2016).
52. Beznosková, P. et al. Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. *PLoS Genet.* **9**, e1003962 (2013).
53. Kouba, T. et al. Small ribosomal protein RPS0 stimulates translation initiation by mediating 40S-binding of eIF3 via its direct contact with the eIF3a/TIF32 subunit. *PLoS ONE* **7**, e40464 (2012).
54. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
55. Nenarokova, A., Záhonová, K. & Nenarokov, S. The high-throughput sequencing datasets. *figshare* https://doi.org/10.6084/m9.figshare.21401541 (2022).
56. Nenarokova, A., Záhonová, K. & Nenarokov, S. Additional data and analyses. *figshare* https://figshare.com/projects/tRNA_anticodon_stem_length_variations_are_critical_for_stop_codon_reassignment/129167 (2022).
57. Nenarokov, S. & Nenarokova, A. Seraff/blasto: annotator & utilities for *Blastocrithidia* project (v1.0.2). *Zenodo* https://doi.org/10.5281/zenodo.7116082 (2022).
58. Potěšil, D. MS analysis of *B. nonstop* proteins. *figshare* https://doi.org/10.6084/m9.figshare.20105417.v2 (2022).

Acknowledgements We thank M. Tesařová (Institute of Parasitology) for help with electron microscopy, O. Namy (Université Paris-Sud) for providing the eRF1 mutant strains, and D. Potěšil and Z. Zdráhal (CEITEC) for help with MS analyses. This work was supported by the Czech Science Foundation grants 18-15962S and 22-14356S (to J.L. and V.Y.), 20-11585S (to Z. Paris) and 20-00579S (to L.S.V.), the Charles University Grant Agency project GAUK 1192819 (to Z. Pavlíková), the Czech Ministry of Education ERD Funds 16_0000759 (to V.Y., Z. Paris and J.L.), the Gordon and Betty Moore Foundation GBMF no. 9354 (to J.L.) and the Praemium Academiae grant provided by the Czech Academy of Sciences (to L.S.V.). Supporting services were supplied by the project LM2018140, ERD Funds 18_046/0015974 and Czech BioImaging LM2018129.

Author contributions A.N., E.H., Z. Paris, L.S.V. and J.L. conceived the study and designed the research. A.K., Z. Pavlíková, A.N., J.V., A.R., P.M., K.Z., S.N., I.M.D., E.H., R.L.R., V.Y., P.B. and Z. Paris carried out the experiments and analysed the data. A.N., Z. Pavlíková, Z. Paris, L.S.V. and J.L. wrote the manuscript. L.S.V. served as the lead contact, completed and submitted the work, and handled the revision and production tasks.

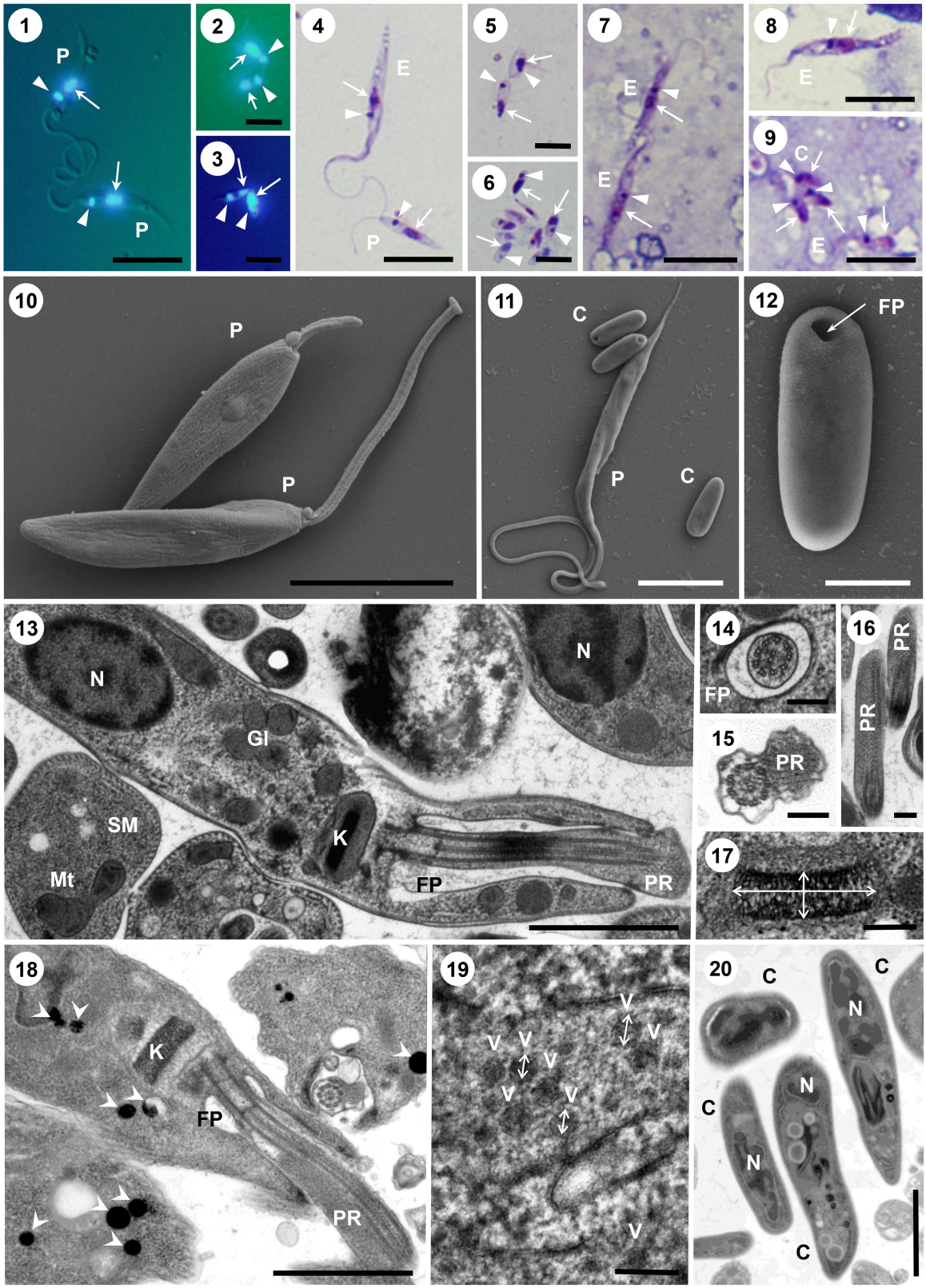
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05584-2.

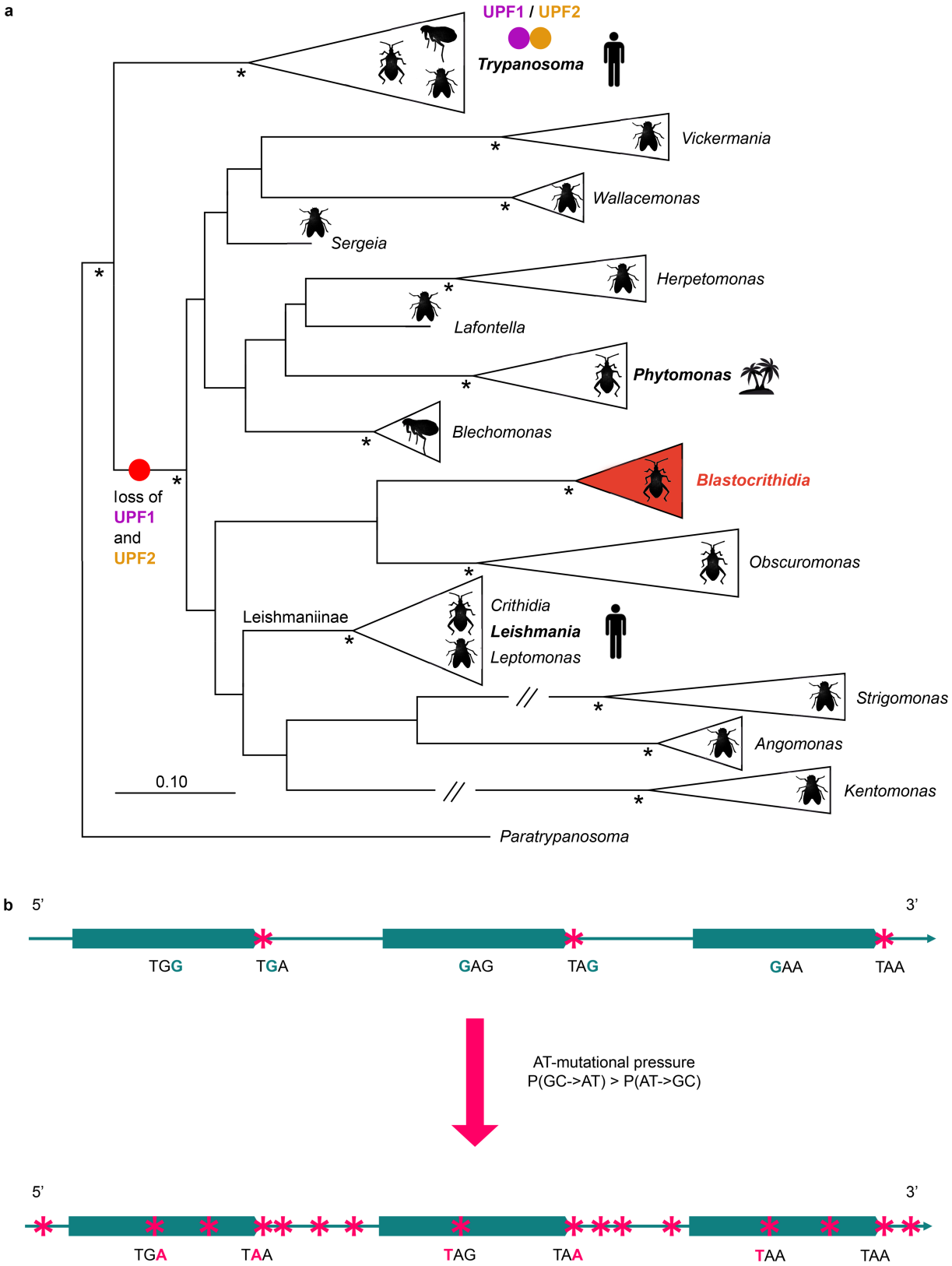
Correspondence and requests for materials should be addressed to Zdeněk Paris, Leoš Shivaya Valášek or Julius Lukeš.

Peer review information *Nature* thanks Pavel Baranov, Olivier Namy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



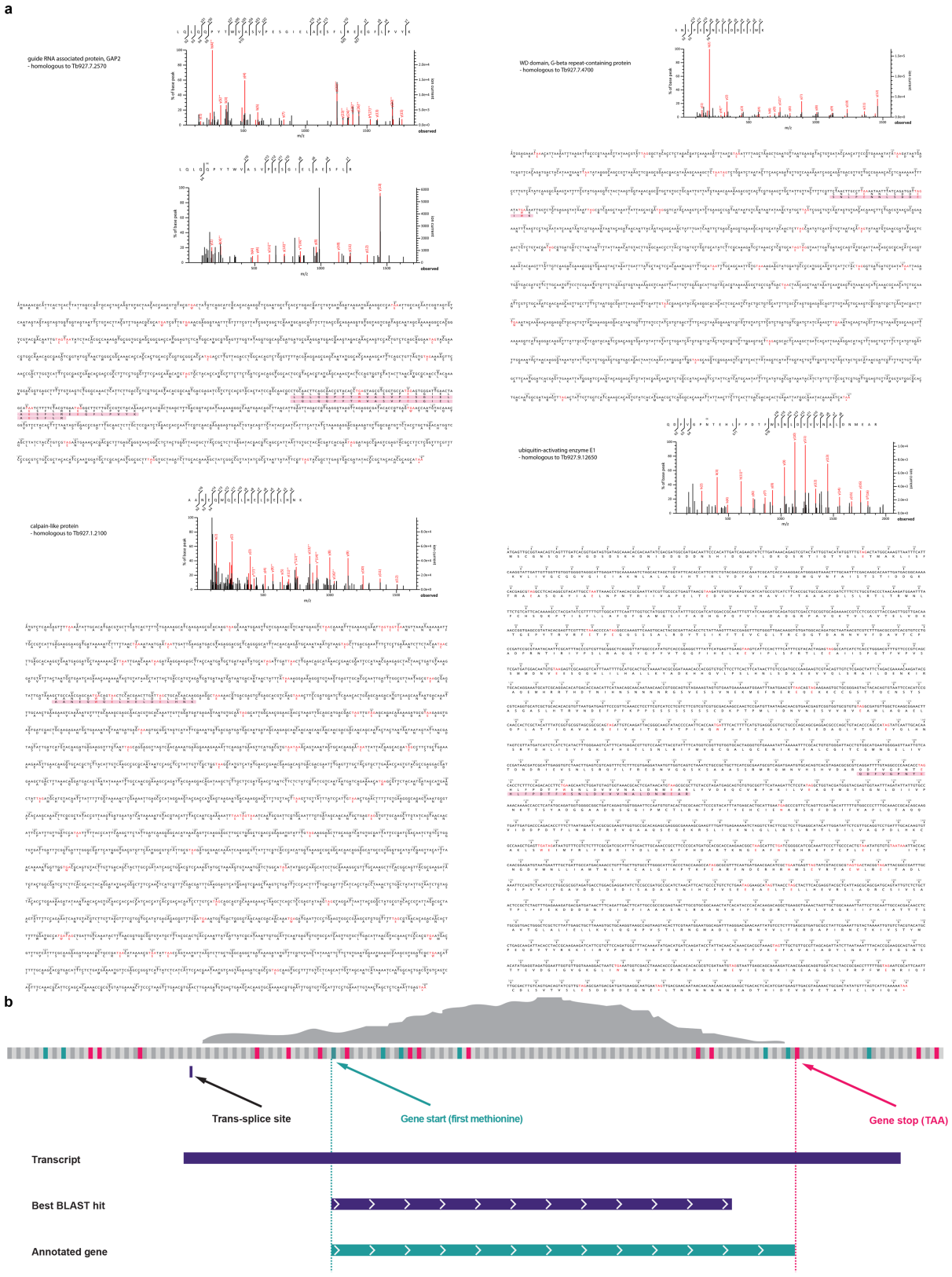
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Morphology of *Blastocrithidia nonstop* sp. nov. Light microscopy (1–9), scanning (10–12) and transmission electron microscopy (13–20). The DAPI-stained (1–3) and Giemsa-stained (4–9) cultured cells (1–6) and parasites from the host insect hindgut (7–9). Promastigote flagellates (P) dominate in the culture (1, 4, 10–11), while the epimastigotes (E) are more frequent in the insect gut (7–9), although they are also present in the culture (4). The cyst-like amastigotes (C) are present in both the culture and host-derived smears (9). The nucleus and the kinetoplast DNA are labelled with arrow and arrowhead, respectively. The scanning electron micrographs show a highly variable length of the flagellum in promastigotes (10) and the cyst-like amastigotes (11 and 12). The longitudinal section thru the cell reveals a deep flagellar pocket (FP; 13 and 18), oval nucleus (N), elongated mitochondria (Mt), oval glycosomes (Gl), and numerous dense acidocalcisomes (arrowhead; 18). The cross- (15) and longitudinally-sectioned (16) external flagellum is equipped with a prominent paraflagellar rod (PR; see also 13 and 18), which is absent in the part of the flagellum within the flagellar pocket (14; see also 13 and 18). Note a thin and wide kinetoplast DNA disk (K; 13, 17, and 18), virus-like particles within the cytoplasm (V; 19), the regularly spaced subpellicular microtubules (SM; 13), and the cyst-like amastigotes (20). Scale bars: 10 µm (1, 4, 7–9, and 10), 5 µm (2, 3, 5, 6, and 11), 1 µm (12, 13, 18, and 20), 200 nm (14–17), and 100 nm (19); repeated 3-times with similar results.



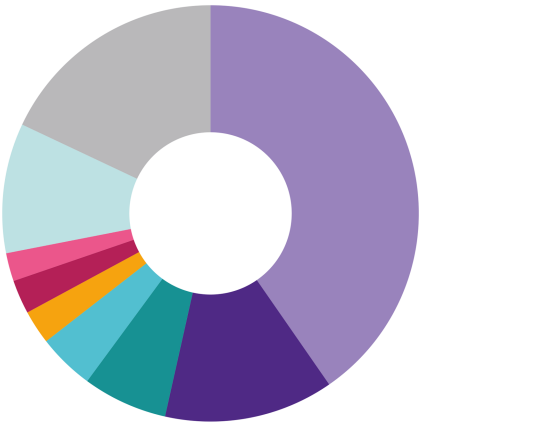
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | The 18S rRNA phylogenetic tree and consequences of AT mutational shift in *Blastocrithidia* ancestor. (**a**) The 18S rRNA-based maximum likelihood phylogenetic tree of the family Trypanosomatidae. The branches are collapsed at the generic level to highlight the inter-generic relationships. The tree was rooted with *Paratrypanosoma confusum*; asterisks mark branches with maximal statistical support (bootstrap values for maximum likelihood >90; Bayesian posterior probabilities >0.95); double-crossed branches are 50% of the original length; the scale bar denotes the number of substitutions per site. The subfamily Leishmaniinae is subdivided into the genera *Crithidia*, *Leishmania sensu lato*, and *Leptomonas*. The genus *Blastocrithidia* is highlighted in red, the principal insect host and/or vectors (fleas, flies, and true bugs) are depicted by pictograms, and three dixenous genera infecting humans or plants are in bold. The presence of the key components (UPF1 and UPF2) of the nonsense-mediated decay pathway, which eliminates mRNAs with premature stops, is mapped onto the tree. The pathway was lost by the predecessor of the trypanosomatid lineage after the separation of the genus *Trypanosoma*. (**b**) Scheme illustrating possible consequences of AT mutational shift in *Blastocrithidia* ancestor. Frequent GC-to-AT substitutions in *Blastocrithidia* ancestor could have led to the following consequences that we observe in *B. nonstop* genome: overall AT-rich genome; extreme AT-richness of intergenic regions with frequent TAA codons; appearance of in-frame stop codons. Possible UAG-to-UAA and UGA-to-UAA substitutions are evolutionary neutral, but allowed TGG-to-TGA, GAG-to-TAG, and GAA-to-TAA substitutions.



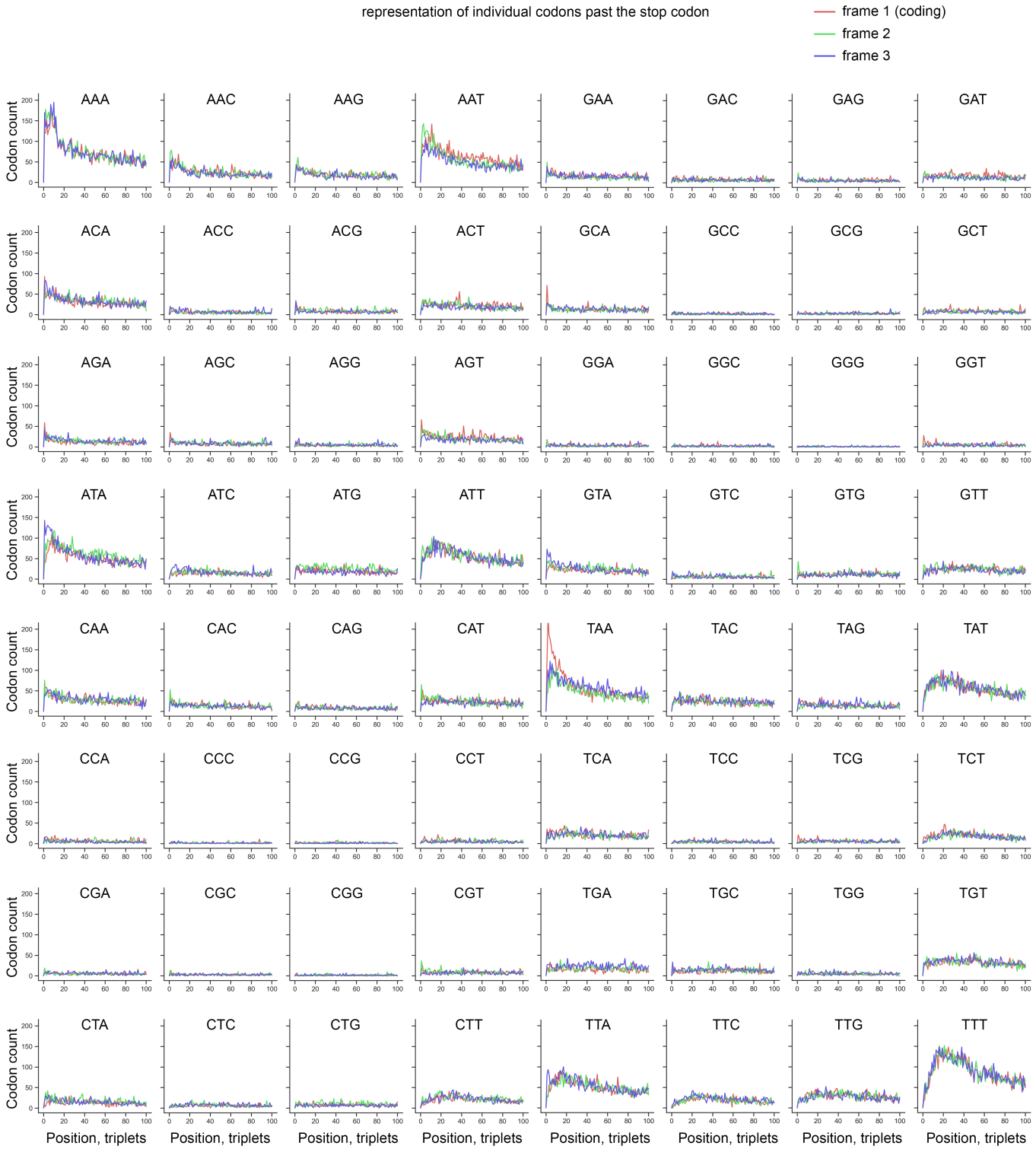
Extended Data Fig. 3 | Proteomic analysis and principles of the structural annotation of *B. nonstop*. (**a**) Proteomic proof of the stop codon reassignment in *B. nonstop*. Four protein-coding genes were selected as they contain all three in-frame stop codons covered by identified peptides from mass spectrometry (MS) analysis. Nucleotide sequences with stop codons in red and their conceptual translations are shown. Below that, identified peptides are highlighted in light-pink. Above the sequences, MS spectra of identified peptide(s) are shown. All identified peptides are available on figshare[58]. (**b**) Principles of the structural annotation of *B. nonstop*. The genes in *B. nonstop* were predicted based on the combination of evidence inferred from transcriptome coverage, *trans*-splicing sites, and the alignments of the reference trypanosomatid proteins. For the detailed protocols, see Methods.

a

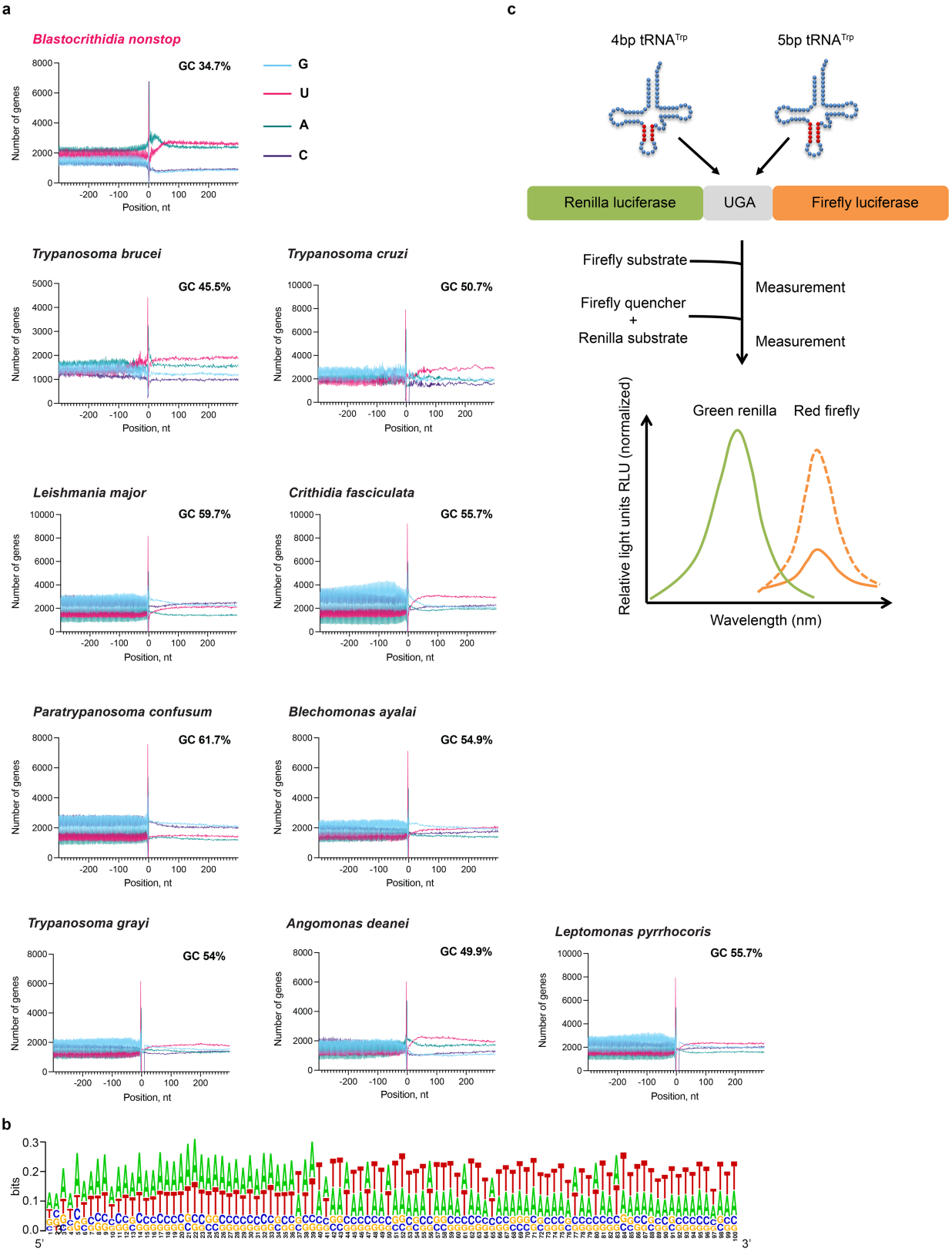


b

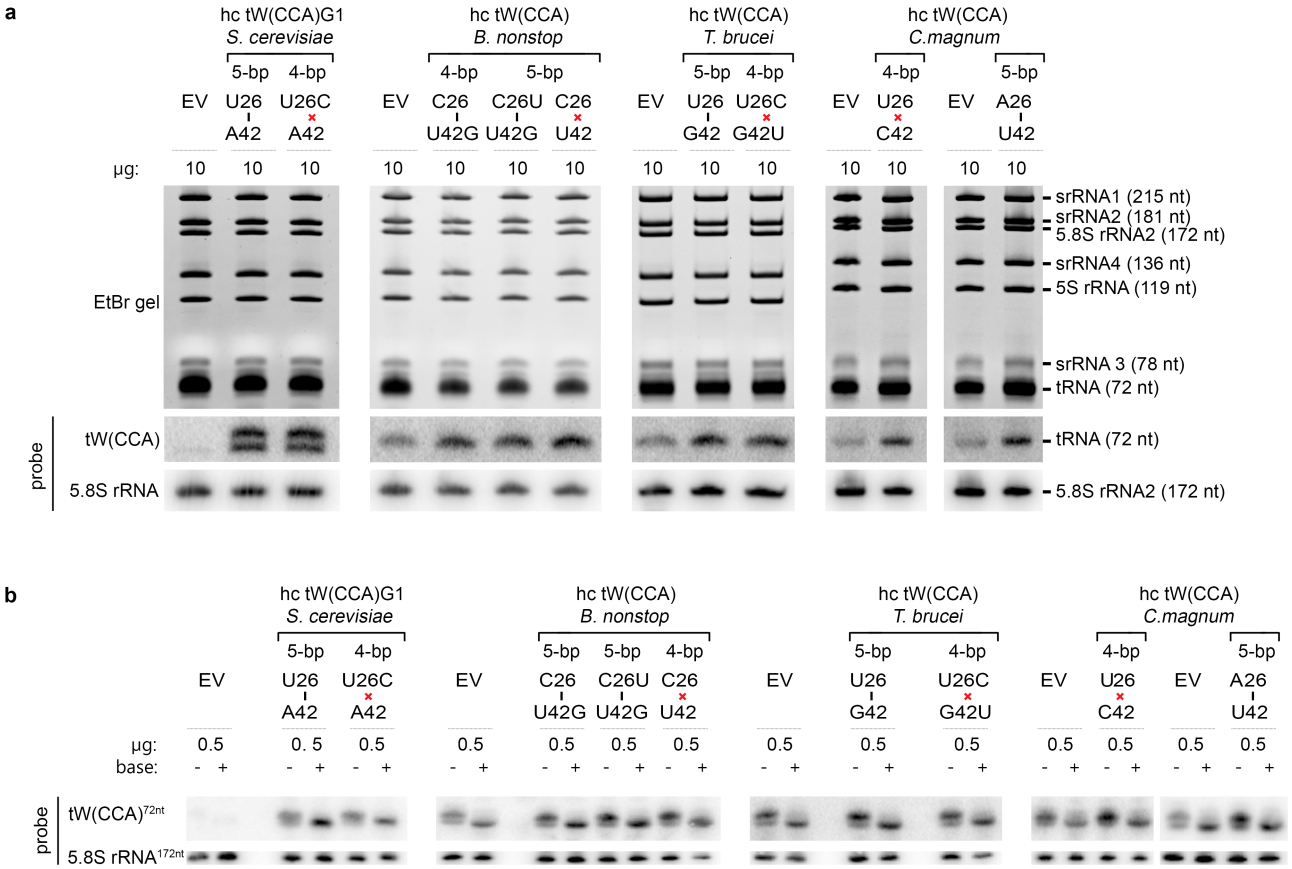
Extended Data Fig. 4 | GO-enrichment analysis and *Blastocrithidia* transcripts with UAA genuine stop codons. (**a**) GO-enrichment analysis of genes devoid of reassigned stop codons. The pie chart summarizes the predicted functions of 228 proteins of *B. nonstop* without reassigned stop codons. The complete list of corresponding proteins with annotations is available in **Source Data** Extended Data Fig. 4. (**b**) UAA is the only genuine stop codon in transcripts from *Blastocrithidia* sp. ex *Lygus hesperus*. Shown are open reading frames from *Blastocrithidia* sp. ex *L. hesperus* that were previously predicted with premature ends[1]. The alignment with *B. nonstop* revealed that all of them can be extended until the next UAA stop codon. Nucleotide identities are highlighted by white font on black background.



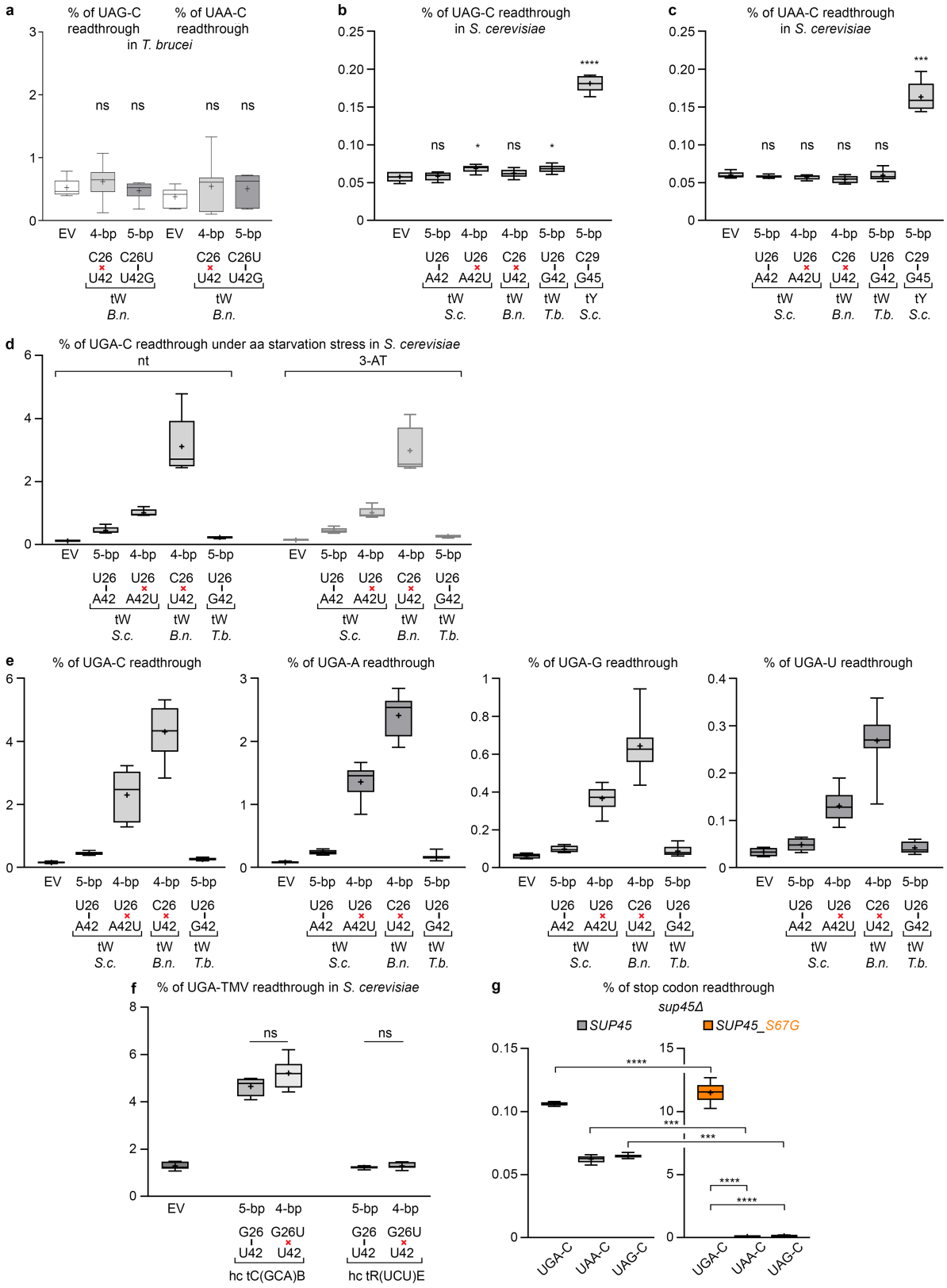
Extended Data Fig. 5 | Distribution of codons in 3′ UTRs of *B. nonstop* transcripts in three coding frames. The graph shows the summary counts for each of 64 codon triplets in 100 triplets (300 nt) after the stop codon in 1,569 *B. nonstop* transcripts calculated in three coding frames. Frame 1 corresponds to the gene coding frame, while frames 2 and 3 are shifted by 1 and 2 nucleotides, respectively.



Extended Data Fig. 6 | Stop codon context analysis and schematic of the dual luciferase assay. (**a**) GC-content around the termination codon position in the 3′ends of protein-coding genes in trypanosomatids. The GC distribution in the *B. nonstop* genome dramatically differs inside of the coding regions *versus* the intergenic regions, in contrast to other trypanosomatids. The former are more GC-rich, however, immediately after the 3′end of genes the genome sequences become substantially more A+T-rich. (**b**) Sequence logo for first 100 nt after the stop codon. A is the most abundant nucleotide in the first ~39 nt past the stop codon with T taking over afterwards. Equiprobable abundance of As and Ts in the first 40 nucleotides after stop codon as a null hypothesis was rejected with the two-tailed p-value <0.0001; $\chi^2 1 = 1218.98$. (**c**) Schematic representation of the dual luciferase assay. A cassette composed of an in-frame UGA codon positioned between the *Firefly* and *Renilla* luciferase genes was electroporated into *T. brucei* or *S. cerevisiae* expressing either 4- or 5-bp-long variants of $tRNA^{Trp}_{CCA}$.

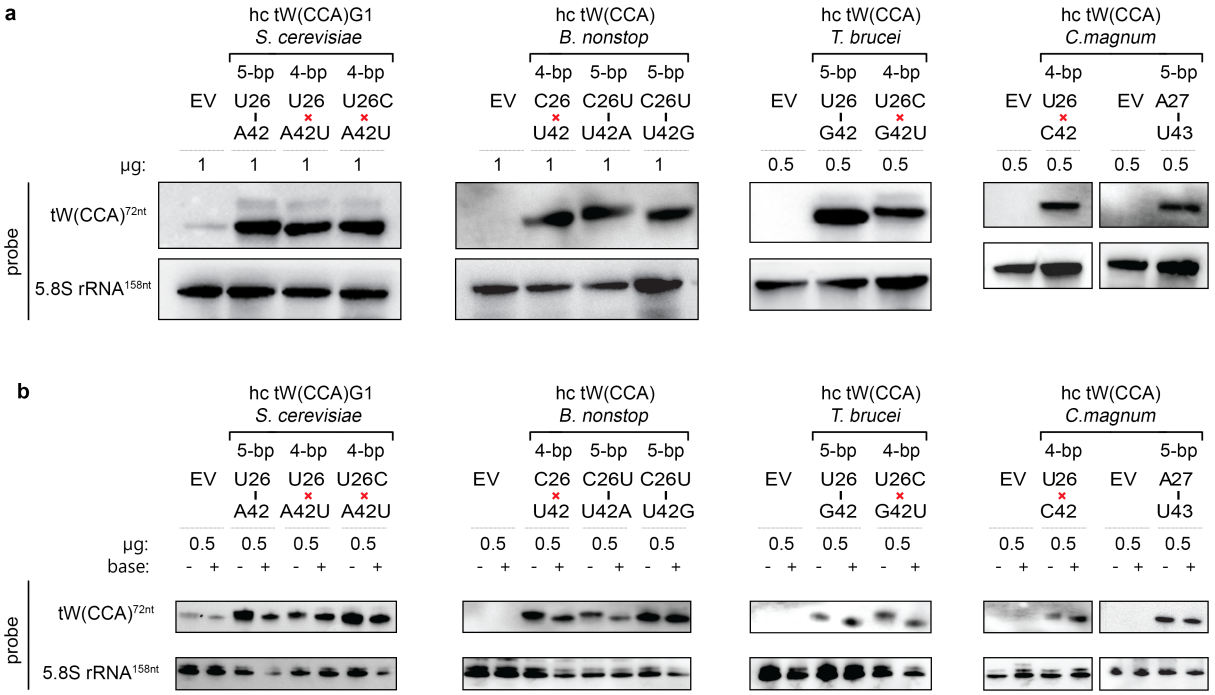


Extended Data Fig. 7 | Detection and charging of tRNA$^{Trp}_{CCA}$ variants expressed in *T. brucei*. (**a**) Increased gene dosage of all analyzed tRNA^{Trp} variants substantially increases their cellular levels in vivo. Total RNA was extracted from *T. brucei* strains expressing plasmid-borne high copy 4-bp vs. 5-bp-long tRNA^{Trp} (hc tW) variants from *S. cerevisiae, B. nonstop, T. brucei*, and *C. magnum* (indicated at the top of each panel). Ten µg of RNA was resolved by urea-PAGE, followed by the northern blot analysis using ^{32}P-labeled oligonucleotides specific for tRNA variants; 5.8S rRNA was used as a loading control. Repeated 3 times with similar results; for gel source data, see Supplementary Fig. 1. (**b**) All expressed tRNA variants show an equal level of aminoacylation. Samples were prepared as in panel a, except that acidic conditions were applied to protect charged tRNAs. Aliquots of RNA samples were subjected to deacylation (labelled as "base +"). Ten µg of acylated or deacylated RNA samples were resolved on the acid-urea 12% polyacrylamide gel, transferred onto a nylon membrane and hybridized with ^{32}P-labeled oligonucleotides specific for given tRNA variants; 5.8S rRNA was used as a loading control. The nature of hydrogen bonding between bases of the anticodon stem 5th base pair (26:42) is indicated by a vertical black line (base paring) or a red cross (no paring). Repeated 3 times with similar results; for gel source data, see Supplementary Fig. 1, for quantifications, see **Source Data** Extended Data Figs. 7 and 9.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Control stop codon readthrough analysis. (**a-c**) The "anticodon stem length" effect is specific only for UGA. (**a**) *T. brucei* cells were transformed with empty vector (EV) or $tRNA^{Trp}_{CCA}$ (tW) with 4-bp or 5-bp AS versions from *B. nonstop* and processed for UAG or UAA readthrough measurements as described in Methods. Readthrough values were normalized to the control cell line (containing dual-luciferase cassette without in-frame stops). The nature of hydrogen bonding between bases of the anticodon stem 5^{th} base pair (26:42) is indicated by a vertical black line (base paring) or a red cross (no paring). Each box in the box-and-whisker plot is represented by n = 9 readthrough values (3 individual experiments each including n = 3 biological replicates). The mean value is marked (+); whiskers range from minimal to maximal values. Statistical significance was determined by the unpaired, two-tailed Welch's *t* test; **** $p < 0.0001$; *** $p < 0.001$; * $p < 0.05$; ns = non-significant. (**b-c**) The yeast strain H541 was transformed with a corresponding dual luciferase readthrough reporter construct YEp-R/T-UAGC-L (panel b) or YEp-R/T-UAAC-L (panel c) together with EV or a given high copy $tRNA^{Trp}$ variant or a control, readthrough-inducing yeast $tRNA^{Tyr}$ (tY). The resulting transformants were grown in synthetic media, processed for stop codon readthrough measurements as described in Methods, and analyzed and plotted as in panel a; **** $p < 0.0001$; *** $p = 0.0002$; * $p < 0.03$; n = 15 values (3 individual experiments each including 5 biological replicates; see also raw data of *Firefly* and *Renilla* measurements in **Source Data** Figs. 3, 4 and Extended Data Fig. 8). (**d**) The "anticodon stem length" effect is independent of amino acid starvation stress. The yeast strain ZH252 was transformed with a corresponding dual luciferase readthrough reporter construct YEp-R/T-UGAC-L together with an EV or a given high copy $tRNA^{Trp}$ variant. The resulting transformants were grown in synthetic media to an O.D. of ~1 and either treated with 3-AT (3-amino-1,2,4-triazole; final concentration 10 mmol/l) for 6 h or not (nt). Subsequently, both cultures were processed for stop codon readthrough measurements as described in Methods, and analyzed and plotted as in panel a; n = 10 values (2 individual experiments each including 5 biological replicates; see also raw data of *Firefly* and *Renilla* measurements in **Source Data** Figs. 3, 4 and Extended Data Fig. 8). (**e**) The nature of the UGA stop codon tetranucleotide does not impact on the difference between readthrough levels displayed by 4-bp-long *vs*. 5-bp-long anticodon stem $tRNA^{Trp}$ variants from various eukaryotes. The yeast strain H541 was transformed with a corresponding dual luciferase readthrough reporter construct (from left to right: YEp-R/T-UGAC-L; YEp-R/T-UGAA-L; YEp-R/T-UGAG-L; YEp-R/T-UGAU-L or YEp-R/T-CAAC-L) together with EV or a given high copy $tRNA^{Trp}$ variant. The resulting transformants were grown in synthetic media and processed for stop codon readthrough measurements as described in Methods, and analyzed and plotted as in panel a; n = 9 values (3 individual experiments each including 3 biological replicates; see also raw data of *Firefly* and *Renilla* measurements in **Source Data** Figs. 3, 4 and Extended Data Fig. 8). (**f**) Shortening of the anticodon stem of *S. cerevisiae* Cys and Arg tRNAs from 5 bp to 4 bp has no effect on UGA-TMV readthrough levels in yeast. The yeast strain H541 was transformed with a corresponding dual luciferase readthrough reporter construct (YEp-R/T-UGA-TMV-L or YEp-R/T-CAAC-L) and subsequently with EV or a given high copy of either Cys or Arg tRNA variant. The resulting transformants were grown in synthetic media, processed for stop codon readthrough measurements as described in Methods, and analyzed and plotted as in panel a; n = 12 values (3 individual experiments, including 3, 4 and 5 biological replicates, respectively; see also raw data of *Firefly* and *Renilla* measurements in **Source Data** Figs. 3, 4 and Extended Data Fig. 8). (**g**) The $sup45^{S67G}$ substitution of yeast eRF1 strictly restricts UGA decoding. The yeast strains bearing wild type eRF1 (*sup45Δ* + *SUP45*) or the *S67G* eRF1 substitution (*sup45Δ* + $sup45^{S67G}$) were transformed with a corresponding dual luciferase readthrough reporter construct (YEp-R/T-UGAC-L; YEp-R/T-UAAC-L or YEp-R/T-UAG-L). The resulting transformants were grown in synthetic media, processed for stop codon readthrough measurements as described in Methods, and analyzed and plotted as in panel a; n = 11 values (3 individual experiments, including 3, 4 and 4 biological replicates, respectively; see also raw data of *Firefly* and *Renilla* measurements in **Source Data** Figs. 3, 4 and Extended Data Fig. 8). Statistical significance was determined by the unpaired, two-tailed Welch's *t* test; **** $p < 0.0001$; *** $p < 0.001$.



Extended Data Fig. 9 | Expression levels and charging efficiency of all tRNAs tested in *S. cerevisiae* in this study. (**a**) Increased gene dosage of all analyzed $tRNA^{Trp}$ variants substantially increases their cellular levels *in vivo*. Total RNA samples were isolated from the yeast strain H541 expressing plasmid-borne high copy 4-bp *vs*. 5-bp-long $tRNA^{Trp}$ (tW) variants from *S. cerevisiae*, *B. nonstop*, *T. brucei*, or *C. magnum* (indicated at the top of each panel). 0.5 or 1 µg of total RNA were resolved on the Criterion Precast gel, transferred onto a nylon membrane and hybridized with a particular DIG-labeled probe against $tRNA^{Trp}$. The 5.8S rRNA was used as a loading control. Repeated 3 times with similar results; for gel source data, see Supplementary Fig. 1. (**b**) Increased gene dosage of all analyzed $tRNA^{Trp}$ variants (4 bp *vs*. 5 bp) shows an equal level of aminoacylation. Total RNA samples were prepared as in panel a, except that acidic conditions were applied to protect charged tRNAs. Aliquots of RNA samples were subjected to deacylation (labelled as "base +"). 0.5 µg of acylated or deacylated RNA samples were resolved on the acid-urea 12% polyacrylamide gel, transferred onto a nylon membrane and hybridized with a particular DIG-labeled probe against $tRNA^{Trp}$. The 5.8S rRNA was used as a loading control. The nature of hydrogen bonding between bases of the anticodon stem 5^{th} base pair (26:42) is indicated by a vertical black line (base paring) or a red cross (no paring). Repeated 3 times with similar results; for gel source data, see Supplementary Fig. 1, for quantifications, see **Source Data** Extended Data Figs. 7 and 9.

Extended Data Fig. 10 | Alignment of eRF1 from various species. Species with non-canonical genetic codes are highlighted by a grey background. X in the sequences represents in-frame stop codons. Blue and yellow boxed alignment regions correspond to the conserved motifs required to recognize stop codons by eRF1. All available *Blastocrithidia* spp. exhibit critical substitution Ser70Gly (numbering according to the human sequence; Ser67 in *S. cerevisiae* or Gly74 in *B. nonstop*). All other motifs are unchanged.

nature portfolio

Corresponding author(s): Leos Valasek, Julius Lukes, Zdenek Paris

Last updated by author(s): Jan 11, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom computer code that was wrote for this project is publicly available on Zenodo (doi:https://doi.org/10.5281/zenodo.7116082)
Data analysis	- FastQC v0.11.5 - BBDuk from BBMap package v36.59 - Ra assembler v0.2.1 - Pilon v1.23 - QUAST v4.4 - Bowtie2 v2.3.4.3 - SAMtools package v1.9 - Cufflinks v2.2.1 - SLaP mapper version Jun-2017 - Trinity v2.3.2 - tRNAScan-SE v2.0.5 - Aragorn v1.2.38 - VARNA v3.93 - MAFFT v7.458 - IQTREE v1.4.1 - vsearch v2.5.2 - BLAST v2.3.0+ - Proteome Discoverer v1.4 - Matrixscience v2.6 - BEDtools v2.26.0 - Marvin v17.3.13.0

- Xcalibur v4.0
- Compound Discoverer v3.0
- mzVault v2.1
- BUSCO v5.2.2
- Prism GraphPad v9.2.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data generated during this study are included in this published article (and its Supplementary Information files). The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE55 partner repository with the dataset identifier PXD033324. The high-throughput sequencing datasets were deposited in the NCBI under the number PRJNA790628 and in the Figshare (https://doi.org/10.6084/m9.figshare.21401541). Additional data and analyses are available in the Figshare (https://figshare.com/projects/tRNA_anticodon_stem_length_variations_are_critical_for_stop_codon_reassignment/129167).

Databases used in this study:
- NCBI - sequencing data (BioProject number PRJNA790628)
- PRIDE repository - mass-spectrometry data (identifier PXD033324)
- Zenodo - computer code (https://doi.org/10.5281/zenodo.7116082)
- Figshare - additional data and analyses (https://figshare.com/projects/tRNA_anticodon_stem_length_variations_are_critical_for_stop_codon_reassignment/129167)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For each experiment we performed at least three biological replicates, which is the standard in the field. The number of repeats N is defined in the main text of the manuscript.
Data exclusions	no data was excluded
Replication	all findings are easily reproducible and were replicated at least 3 times
Randomization	this is not relevant because no experimental groups of organisms were generated in our study
Blinding	All the experiments did not involve any binding measurements, because it is not required for these types of analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

*anti-renilla monoclonal antibody MAB4410-1 clone 1 D5.2 Lot. No. Q3666887 (Merck END Millipore); used at 1:500
*mouse Anti-Tubulin monoclonal antibody A11126 (Thermo Fischer Scientific); used at 1:7000
*anti-mouse HRP A9044 produced in rabbit Lot. No. 089M4797V (Sigma); used 1:2000
*anti-eRF1/Sup45 polyclonal rabbit antibody - prepared in house in co-operation with the commercial subject Appronex in 2013 (Prague, Czech Republic); used at 1:500
*anti-Rps0A polyclonal rabbit antibody - prepared in house in co-operation with the commercial subject Appronex in 2012 (Prague, Czech Republic); used at 1:500
*ECL- anti rabbit IgG, Horseradish Peroxidase, linked whole antibody from donkey, Lot17010251
https://www.sigmaaldrich.com/CZ/en/product/sigma/gena934100ul; used at 1:10000

Validation

Preparation of non-commercial antibodies against eRF1/Sup45
The GST-SUP45 fusion protein encoded by pGEX-SUP45 was expressed in E. coli and purified from the WCE by incubation with Glutathione-Sepharose 4B beads (Pharmacia). The isolated protein was resolved by SDS-PAGE (4–20% gels), excised from the gel, and washed with 1× PBS. Altohgether 10 rabbits, whose pre-immune sera were collected for specificity testing, were injected with the purified protein and sera containing polyclonal antibodies against Sup45 were obtained commercially by Apronex (Prague, the Czech Republic). The specificity of antibodies was tested with pre-immune sera as a control in our lab and the best scoring sera (with the least cross-reactions) were selected for scientific purposes.
Beznosková, P. et al. Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. PLoS Genet 9, e1003962 (2013).
Dilution: 1:500

Preparation of non-commercial antibodies against Rps0A
The GST-RPS0A fusion protein encoded by pGEX-RPS0A was expressed in E. coli and purified from the WCE by incubation with Glutathione-Sepharose 4B beads (Pharmacia). The isolated protein was resolved by SDS-PAGE (4–20% gels), excised from the gel, and washed with 1× PBS. Altohgether 10 rabbits, whose pre-immune sera were collected for specificity testing, were injected with the purified protein and sera containing polyclonal antibodies against Rps0awere obtained commercially by Apronex (Prague, the Czech Republic). The specificity of antibodies was tested with pre-immune sera as a control in our lab and the best scoring sera (with the least cross-reactions) were selected for scientific purposes.
Zdroj : Kouba T. et al. (2012) Small ribosomal protein RPS0 stimulates translation initiation by mediating 40S-binding of eIF3 via its direct contact with the eIF3a/TIF32 subunit. PLoS One 7, e40464.
Dilution: 1:500