

# SCIENTIFIC REPORTS



OPEN

## Large-scale mass spectrometry-based analysis of *Euplotes octocarinatus* supports the high frequency of +1 programmed ribosomal frameshift

Ruanlin Wang, Zhiyun Zhang, Jun Du, Yuejun Fu &amp; Aihua Liang

Programmed ribosomal frameshifting (PRF) is commonly used to express many viral and some cellular genes. We conducted a genome-wide investigation of +1 PRF in ciliate *Euplotes octocarinatus* through genome and transcriptome sequencing and our results demonstrated that approximately 11.4% of genes require +1 PRF to produce complete gene products. While nucleic acid-based evidence for candidate genes with +1 PRF is strong, only very limited information is available at protein levels to date. In this study, *E. octocarinatus* was subjected to large-scale mass spectrometry-based analysis to verify the high frequency of +1 PRF and 226 +1 PRF gene products were identified. Based on the amino acid sequences of the peptides spanning the frameshift sites, typical frameshift motif AAA-UAR for +1 PRF in *Euplotes* was identified. Our data in this study provide very useful insight into the understanding of the molecular mechanism of +1 PRF.

Protein translation is an essential life process, in which both efficiency and fidelity are important. A ribosome translates the triplet genetic codes on an mRNA into amino acids of a protein in high fidelity<sup>1</sup> which requires the strict maintenance of in-frame reading of the genetic codes of the mRNA. As mRNA shifts, a truncated and non-sense protein will be produced, resulting in an increase in the energetic cost of translation, and additional loads for cellular cleanup and quality control machineries<sup>2</sup>. However, in special cases, the translating ribosome can switch from the initial (0) reading frame to a -1 or +1 reading frame at a specific position, and then continues its translation<sup>3</sup>. This process is called programmed ribosomal frameshifting (PRF). PRF commonly occurs in all life forms, including bacteria and eukaryotes<sup>4-6</sup>.

Researchers have found two main PRFs (-1 and +1) in viruses and other cellular organisms. The -1 PRF occurs frequently in viruses and other cellular organisms with well documented roles. In viruses and bacteria, for example, such frameshifting is used to expand the information content of their genomes<sup>7,8</sup>. Recent results have implicated the roles of -1 PRF in quality control of mRNA and DNA stability in eukaryotes<sup>9,10</sup>. Unlike -1 PRF, +1 PRF are less frequently found in bacteria, fungi, mammals and ciliated protozoa of *Euplotes*. In bacteria, +1 PRF is used as a sensor and effector of an autoregulatory circuit to express release factor 2<sup>11,12</sup>. In mammals and fungi, such frameshifting is widely employed to regulate the expression of ornithine decarboxylase antizyme, a negative regulator of cellular polyamine levels<sup>13,14</sup>. Protein sequencing has also revealed that +1 PRF frameshifting is used to express the *Tsh* gene of *Listeria monocytogenes* phage PSA, the *PA-X* gene of influenza A virus and the *pol* gene of the *Saccharomyces cerevisiae* retrotransposons Ty1 and Ty3<sup>15-18</sup>.

The +1 PRF is more frequently used for *Euplotes* genes than in genes of other known species<sup>19-21</sup>. Our previous survey indicated that approximately 11% of the genes in *Euplotes octocarinatus* require one or more +1 PRF to produce their protein products<sup>19</sup>. In addition, +1 PRF genes have been identified in other *Euplotes* species (Table 1). In all *Euplotes* cases, DNA sequencing has revealed two separate open reading frames (ORFs) that, if joined by translational frameshifting, would produce a single protein similar to homologous ones in other species.

Key Laboratory of Chemical Biology and Molecular Engineering of Ministry of Education, Institute of Biotechnology, Shanxi University, Taiyuan 030006, China. Correspondence and requests for materials should be addressed to A.L. (email: aliang@sxu.edu.cn)

Species	Genes	Putative slippery sequences	Methods & References
<i>Euplotes raikovi</i>	Mitogen-activated protein kinase	AAA UAA A	Western blot <sup>22</sup>
<i>Euplotes nobilii</i>	Mitogen-activated protein kinase	AAA UAA A	Western blot <sup>23</sup>
<i>Euplotes aediculatus</i>	La motif protein	AAA UAA A	Mass spectrometric analysis <sup>24</sup>
<i>Euplotes octocarinatus</i>	Nuclear protein kinase	AAA UAA A	Nucleic-acid-based analysis <sup>37</sup>
	cAMP-dependent protein kinase	AAA UAA A	Nucleic-acid-based analysis <sup>26</sup>
<i>Euplotes minuta</i>	Reverse transcriptase subunits of telomerase	AAA UAA A	Nucleic-acid-based analysis <sup>38</sup>
<i>Euplotes vannus</i>	Reverse transcriptase subunits of telomerase	AAA UAA G	Nucleic-acid-based analysis <sup>38</sup>
<i>Euplotes crassus</i>	Reverse transcriptase subunits of telomerase-1	AAA UAA C & AAA UAA G	Nucleic-acid-based analysis <sup>39</sup>
	Reverse transcriptase subunits of telomerase-2	AAA UAA C	Nucleic-acid-based analysis <sup>27</sup>
	Reverse transcriptase subunits of telomerase-3	AAA UAA C & AAA UAA G	Nucleic-acid-based analysis <sup>27</sup>
	C <sub>2</sub> H <sub>2</sub> -type zinc finger protein	AAA UAA U	Nucleic-acid-based analysis <sup>20</sup>
	MORN repeat protein	AAA UAA G	Nucleic-acid-based analysis <sup>20</sup>
	Ser/Thr protein kinase	AAA UAA C & AAA UAA G & AAA UAA A	Nucleic-acid-based analysis <sup>20</sup>
	Tec2 transposon ORF2	AAA UAG U	Nucleic-acid-based analysis <sup>40</sup>

**Table 1. Summary of +1 PRF genes in ciliated protozoa of the genus *Euplotes*.**

Although nucleic acid-based evidence for frameshifting is strong, information at protein levels is limited (Table 1). For instance, Western blot analysis on the expression of *MAPK1* genes in *E. raikovi*<sup>22</sup> and *E. nobilii*<sup>23</sup> revealed a mass close band of MAPK1 using different antibodies, confirming the presence of +1 PRF. Mass spectrometry (MS) has also been conducted to analyze the La-motif protein associated with telomerase in *E. aediculatus*<sup>24</sup>. Several peptides derived from the purified La-motif protein were sequenced. One peptide was found to be encoded within the 0 frame ORF, while the remaining peptides are encoded by the +1 frame ORF. These data indicate that the protein is produced by frameshifting. However, the precise site of the frameshift in *Euplotes* remains unclear because of the lack of peptides that actually spanning the frameshift site.

In our current study, total proteins of *E. octocarinatus* were subjected to large-scale MS-based analysis through shotgun liquid chromatography tandem MS (LC-MS/MS). A total of 2,842 proteins were detected, among which 226 were translated via +1 PRF. Furthermore, seven frameshift sites in six proteins were covered by one or two unique peptides. The amino acid sequences of these peptides indicated that the frameshift occurred at “U” of the slippery sequence “AAA-UAR” in *Euplotes*. Moreover, 14 +1 PRF proteins with putative novel slippery sequences were detected. These results provided evidence for the reality of these novel slippery sequences in *Euplotes*. Our data shed light onto the molecular mechanism of +1 PRF in *E. octocarinatus*.

## Results

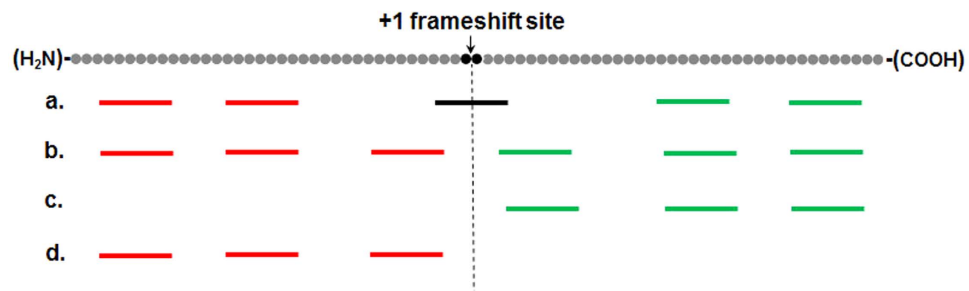
### Large-scale MS-based analysis of *E. octocarinatus* supports the high frequency of +1 PRF.

Previous work suggested that a high frequency of +1 PRF existed in *E. octocarinatus*<sup>19</sup>. A shotgun analysis was applied to investigate the proteome of *E. octocarinatus* to experimentally verify such high frequency +1 PRF *in vivo*. The *Euplotes* protein sample was initially digested with trypsin, which cleaved peptides at the C-terminal side of a lysine (K) and an arginine (R) residues. However, the AAA codon, coding for a lysine, immediately preceded the stop codon of the 0 frame ORF in the majority of +1 PRF genes (approximately 94.2%) of *E. octocarinatus*<sup>19</sup>. Therefore, sample digestion with trypsin likely affects the recovery of the slippery sites. Two additional protein samples were digested with GluC and chymotrypsin to obtain peptide spanning the shift site. A total of 2,842 proteins were obtained from the analyses of the three samples, including 226 +1 PRF proteins, based on the criteria reported in the literature<sup>25</sup>. The previously reported +1 PRF gene in *E. octocarinatus*, namely, cAMP-dependent protein kinase<sup>26</sup> (CUFF.28794.1), was also identified in this study. Two separate peptides of cAMP-dependent protein kinase encoded by the +1 frame ORF were obtained (see supporting information). The peptide sequences of 226 +1 PRF proteins identified by LC-MS/MS were presented in the supporting information.

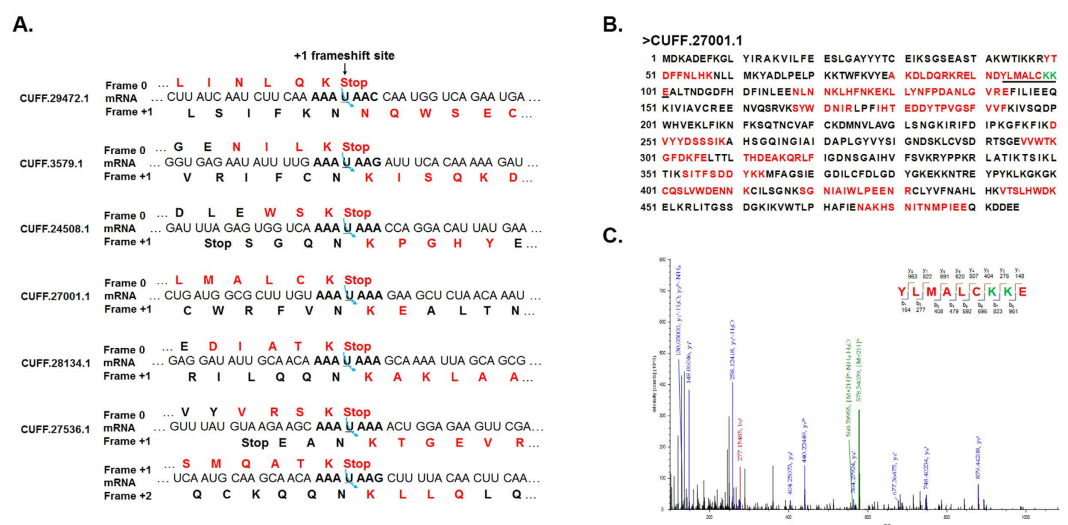
The MS data provided direct or indirect protein evidence for the translation of these +1 PRF candidate genes requiring frameshifts *in vivo*. The observed high frequency of approximately 8% (226 of 2,842 proteins) of frameshifting also supported the notion that euplotids possess an extremely high number of +1 PRF genes.

**PRF of *E. octocarinatus* occurs within the motif “AAA-UAR” in the +1 direction.** Based on their positions of detected peptides relative to the potential frameshift site, 226 +1 PRF proteins were divided into four classes (Fig. 1): (a) the frameshift site was covered by one or two peptides (6 proteins); (b) both the upstream (0 frame ORF) and the downstream (+1 frame ORF) of the frameshift site was covered by peptides (81 proteins); (c) only the downstream (74 proteins) or (d) the upstream (65 proteins) of the frameshift site was covered by peptides (full output of the MS/MS analysis is presented in Table S1 and Table S2.)

We focused on the six proteins whose frameshift sites were covered. Except for CUFF.29472.1 whose frameshift site was covered by two peptides (Fig. S2), the frameshift sites of the other five proteins were covered by a unique peptide (Figs 2B and 3A and Figs S3–S5). The amino acid sequence of these peptides allowed the determination of the location and direction of the shift. Figure 2A shows the nucleotide sequences in the vicinity of the seven frameshift sites (CUFF.27536.1 has two frameshift sites) along with the predicted translation products from the 0



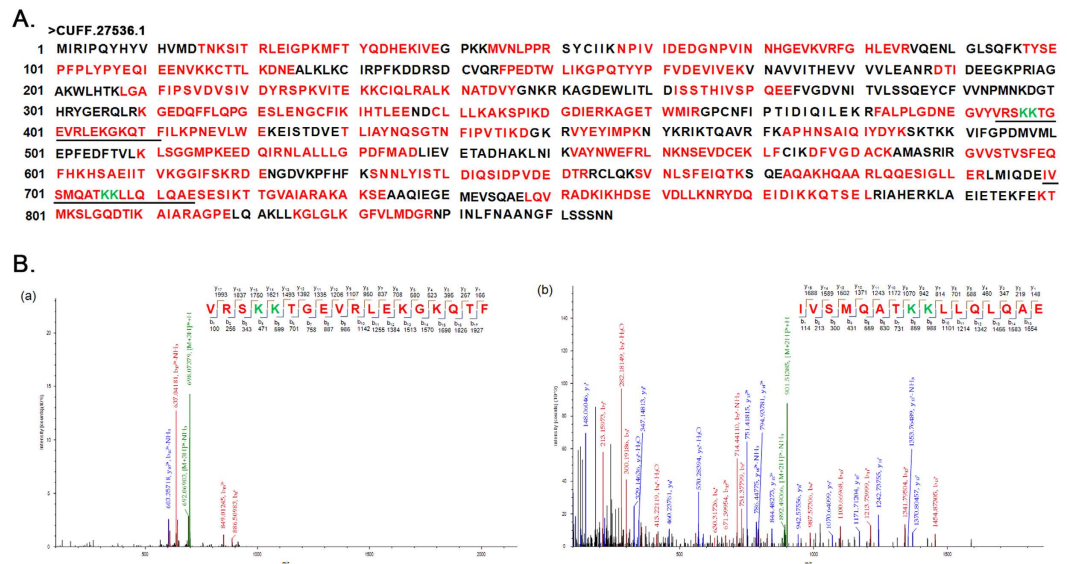
**Figure 1. Overview of the LC-MS/MS result of +1 PRF proteins.** The dots on top of the figure indicate the amino acids of the detected +1 PRF proteins. Dark black dots correspond to the putative frameshift site. The lines refer to the peptides identified by MS. Black line shows the peptide that spans the frameshift site. Red lines and green lines indicate peptides located in the upstream and downstream of the frameshift site, respectively. The 226 detected +1 PRF proteins are divided into four classes (a, b, c and d).



**Figure 2. MS analysis of the six frameshift proteins.** (A) Close-up of the frameshift region is shown. The AAA-UAA motif is shown in bold. The +1 frameshift events are illustrated by curved arrows at the “skipped” nucleotides, which are underlined. The conceptual translations in the 0 reading frame and +1 reading frame are aligned above and below the mRNA sequence, respectively. The amino acids of the peptides identified through mass spectrometry are indicated in red. (B) Complete amino acid sequence of the CUFF.27001.1 protein. The peptides identified by MS are indicated in red. The peptide spanning the frameshift site is underlined. The putative frameshift site is highlighted in green. (C) LC-MS/MS fragmentation spectrum of the shift site peptide YLMALCKKE from CUFF.27001.1. The insert shows the peptide sequence with ‘b-’ and ‘y-’ type fragment ions that strongly support the shift site peptide identified in the LC-MS/MS analysis. The protein was alkylated with iodoacetamide to protect Cys residues.

and +1 reading frames. All the seven sequences indicated that a frameshift apparently occurred at the “U” of the “AAA-UAA” when the translating ribosome was transferred from the AAA lysine codon to the next codon (AAA, AAG or AAC). The +1 shifts were also proved by the identification of the specific peptide fragments. Although the possible products of -2 shifts would feature the same C-terminal sequences, the fragments overlapping the slippery sequence would contain an additional amino acid (the first residue was introduced into the chain after shifting back), and would therefore have a mass more than 100 Da difference. However, the determined masses of the individual polypeptide fragments were very close to the calculated masses for a +1 shift. Therefore, +1 shift rather than -2 shifts is used to produce the complete protein products.

**Evidence for the presence of two +1 PRFs in one single gene of *E. octocarinatus*.** Nucleic acid-based analysis suggested that several genes of *Euplotes* may use two<sup>27</sup> or three<sup>20</sup> +1 PRFs to produce their protein products. The CUFF.27536.1 protein was a major vault protein that may require two +1 frameshifts for expression (Fig. 3A). The shotgun LC-MS/MS analysis yielded a total fragment covering 64% of this protein. Two peptides spanning the two putative frameshift sites were identified (Fig. 3B). The peptide VRSKKTGEVRLEKQTF defined the frameshift site (AAA-UAA A) and direction (+1) of the first



**Figure 3.** MS analysis of the CUFF.27536.1 protein. (A) Complete amino acid sequence of the CUFF.27536.1 protein. The peptides identified by MS are indicated in red. The peptide spanning the frameshift site is underlined. The two frameshift sites are highlighted in green. (B) LC-MS/MS fragmentation spectrum of the two shift site peptides “VRSKKTGEVRLKQKQTF” and “IVSMQATKLLQLQAE” from CUFF.27536.1. The insert shows the peptide sequence with “b-” and “y-” type fragment ions that strongly support the shift site peptides identified in the LC-MS/MS analysis.

frameshifting, the peptide IVSMQATKLLQLQAE clearly defined the frameshift site (AAA-UAG) and the direction (+1) of the second frameshifting (Fig. 2A).

The other three proteins (CUFF.16975.1, CUFF.4515.1 and CUFF.7325.1) were also identified to harbor two frameshift sites (see supporting information). In all three cases, the peptides that were encoded by all three reading frames were detected, thereby providing a solid evidence confirming that a single protein is produced by two +1 frameshifting processes.

**Possible novel slippery sequences for +1 PRF.** In addition to the 212 +1 PRF proteins with the classical “*Euplotes* frameshift motif” (5'-AAA-UAR-3'), our data also suggested the presence of novel +1 PRF slippery sequences in other 14 proteins (Table 2). Eight of these proteins showed peptides covering both upstream and downstream of the frameshift site, thereby providing a clear indication that a complete protein was produced via frameshifting.

Due to the inability to obtain peptides spanning the frameshift site, we could not determine their precise frameshift sites. However, the frameshift site could be deduced from further sequence analysis. The CUFF.26295.1 protein is a serine hydroxymethyltransferase (SHMT) with a putative slippery sequence “UUU-UAGA” (Fig. 4). In this protein, an overlap occurred between the two ORFs because the +1 frame ORF has a termination codon located at the 71 bases upstream of the termination codon of the 0 frame ORF (Fig. 4B). This region represented part of the conserved SHMT domain. The alignment of the CUFF.26295.1 protein with SHMT proteins from a set of evolutionarily diverse organisms revealed several highly conserved residues (Fig. 4C), among which, Ser<sub>243</sub> and Tyr<sub>245</sub> could participate in catalysis or stabilizing the structure<sup>28</sup>. Furthermore, Arg<sub>253</sub> was also highly conserved. It was followed by a less conserved phenylalanine or leucine. These data supported the frameshift event that may occur at “UUU-UAGA” motif.

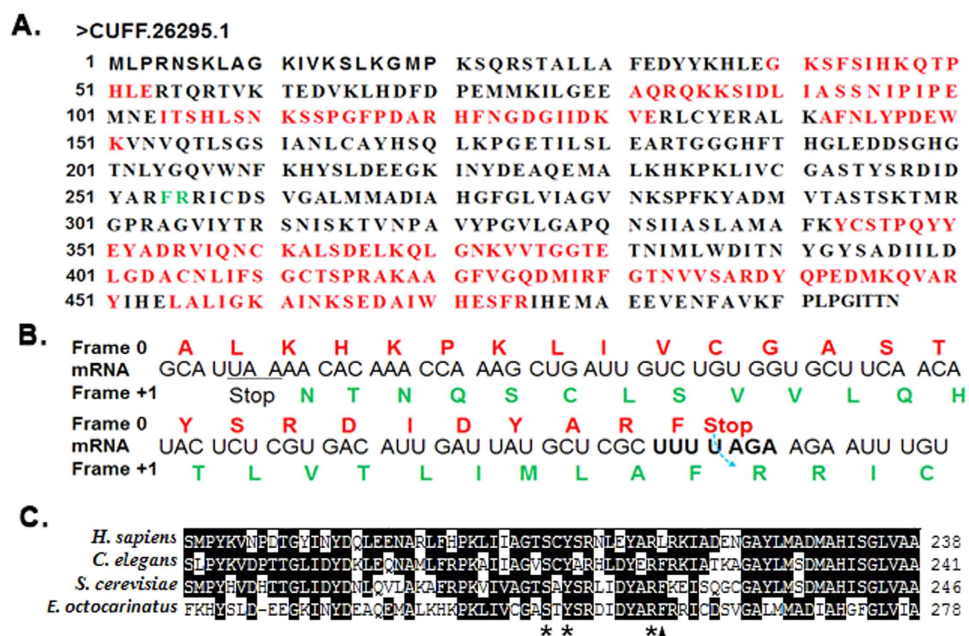
## Discussion

In this article, we performed a large-scale MS-based analysis of *E. octocarinatus*-derived protein samples and observed extremely high frequency of +1 PRF. In total, 226 +1 PRF proteins were detected, of which we obtained seven peptides spanning the frameshift sites. The amino acid sequences of these peptides suggested that the frameshift occurred at the slippery motif “AAA-UAR” in *Euplotes*. One of the six proteins, CUFF.27536.1, provided solid evidence indicating that a single protein was produced by two +1 frameshifting. Furthermore, putative novel slippery motifs were detected in 14 +1 PRF proteins, suggesting the presence of novel +1 PRF proteins in *Euplotes*.

In our previous study, a genome-wide investigation of *E. octocarinatus* based on its genome and transcriptome sequencing indicated that approximately 11% genes required +1 PRF to produce the corresponding protein products<sup>19</sup>. The observed 8% (226 of 2,842 proteins) frequency of frameshifting in this study was somewhat lower than the 11% from the previous survey. This inconsistency may be due to the fact that the majority of +1 PRF genes in *Euplotes* express low-abundant proteins<sup>19,20</sup> that might not be detected in an MS analysis. Nevertheless, our results still supported the notion that euplotids possessed an extremely high number of genes requiring a +1 frameshift for expression. The MS data also showed the presence of two +1 PRFs in one single

Sequence ID	Putative slippery sequence	Unique Peptides
CUFF.26295.1	UUU UAG	16
CUFF.26296.1	UUU UAG	15
CUFF.8736.1	UUU UAA	14
CUFF.25615.1	UUU UAA	3
CUFF.1984.1	UUU UAA	1
CUFF.21695.1	AUU UAA	4
CUFF.2970.1	AUU UAA	1
CUFF.28108.1	AUU UAG	1
CUFF.4268.1	AAG UAA	1
CUFF.166.1	ACC UAA	11
CUFF.13288.1	AGA UAG	2
CUFF.16329.1	UAU UAG	3
CUFF.9615.1	CUU UAA	1
CUFF.2860.1	AAU UAA	1

**Table 2.** +1 PRF proteins with novel slippery sequence.



**Figure 4. Sequence analysis of the CUFF.26295.1 protein.** (A) Complete amino acid sequence of the CUFF.26295.1 protein. The peptides identified by MS are indicated in red. The putative frameshift site is highlighted in green. (B) Close-up of the putative frameshift region is shown. The UUU-UAGA motif is shown in bold. The putative +1 frameshift events are illustrated by curved dashed arrow at the “skipped” nucleotide. Conceptual translations in the 0 reading frame (red) and +1 reading frame (green) are aligned above and below the mRNA sequence, respectively. (C) The parts of the predicted protein sequence of the CUFF.26295.1 protein (*E. octocarinatus*) are aligned with the respective regions of SHMT proteins from humans (*H. sapiens*), *Caenorhabditis elegans* (*C. elegans*), and *Saccharomyces cerevisiae* (*S. cerevisiae*). The amino acids in black shading are identical in all four proteins. The numbers to the right of each of the sequence refer to their end amino acid positions. The sequence shown in this figure assumes the frameshift event at “U” of the “UAGA”. The conserved residues is marked by an asterisk. The putative location of the frameshift is marked by a black triangle.

gene in *E. octocarinatus*, indicating that two frameshifts are required to produce a complete protein. Failure at any frameshift site because of low efficiency will end up with a truncated and/or deleterious protein. The observed high frameshift frequency suggested that eulplotids may possess a unique mechanism to process frameshifts efficiently.

In our previous study, we identified 211 novel +1 PRF genes with different types of slippery sequences<sup>19</sup>. In the present study, the MS data provided protein evidence for eight types of novel slippery sequences (Table 2). However, we were unable to obtain peptides that actually span the frameshift site. Nevertheless, the alignment of our putative proteins sequence with homologous counterparts of a set of evolutionarily diverse organisms

showed the conservation of functionally important amino acid residues. This provided an indirect evidence for the existence of these novel slippery sequences. The diverse slippery sequences complicate our understanding on the mechanism behind +1 frameshift in *Euplotes*, which warrants future study.

Stimulatory elements, such as upstream Shine-Dalgarno-like sequences or downstream pseudoknot structures, have been shown to promote efficient frameshifts in other organisms<sup>1,3</sup>. However, none of these sequences were seen to associate with the frameshift in *Euplotes*. The only common feature was the “shifty stop” (UAA or UAG) in the slippery site. In *Euplotes*, the codon UGA was reassigned as cysteine<sup>29</sup> or selenocysteine<sup>30</sup>. Previous study has shown that the reassignment of UGA to Cys in *E. octocarinatus* results in an increased +1 PRF at both UAA and UAG codons<sup>31</sup>. However, the poor recognition of the terminators was necessary but not sufficient to evolve efficient frameshifting<sup>31</sup>. A possible factor may be the unusual tRNAs in *E. octocarinatus*. Expanded or modification-deficient anticodon stem loops could promote +1 translational frameshifting<sup>32–34</sup>. In our previous study, we reported a suppressor tRNA of UAA with an apparently nine-base anticodon loop<sup>19</sup>. Subsequently, we identified two genes for the AAA decoding tRNA<sup>Lys</sup> with nine-base anticodon loops from the *E. octocarinatus* genome (data not shown). The structure study of the 70S ribosome bound to frameshift suppressor tRNA<sup>SufA6</sup> and N1-methylguanosine at position 37 modification-deficient anticodon stem loop revealed that the disruption of the conserved U32-A38 base pair promotes +1 decoding<sup>34</sup>. The expansion of anticodon loops to nine-base in these unusual tRNAs might also disrupt the interaction of the 32–38 pair, thereby causing the +1 frameshifting in *E. octocarinatus*. Further experimental verification is needed to clarify whether and how these unusual tRNAs regulate +1 PRF in *E. octocarinatus*. A survey of these unusual tRNAs in other species of the genus *Euplotes* is also needed.

An additional issue that need to be addressed is the function of +1 PRF in euplotids. PRF plays a role in regulating gene expression of some other organisms<sup>1,3</sup>. However, there is no experimental evidence showing that +1 PRF has such a role in *Euplotes*. Since *E. octocarinatus* does appear to use frameshift frequently, it would be of significance to further investigate the role of this PRFs. In euplotids, frameshifting invariably results in a substantial extension of translation and accordingly euplotids may have evolved an efficient frameshifting system to process the unusually high numbers of PRFs; as such, the decrease in protein expression can be negligible. In this case, +1 PRF unlikely plays a regulatory role in these organisms. Further studies on the expression of individual genes under different conditions should be performed to clarify this issue.

## Methods

**Sample Preparation.** Cells of line 69 of *E. octocarinatus* were cultured and harvested as described previously<sup>19</sup>. The harvested cells were snap-frozen in liquid nitrogen and then homogenized in SDT buffer (4% SDS, 1 mM DTT, 150 mM Tris-HCl, pH 8). After 15 min incubation in boiling water, the homogenate was subjected to continuous sonication treatment on ice. This crude extract was then clarified via centrifugation at 12,000 g and 4 °C for 15 min. Proteins were precipitated by adding 1/5th volume of 100% (w/v) trichloroacetic acid (TCA) and the protein pellet was collected via centrifugation at 12,000 g and 4 °C for 30 min and incubated overnight in 1 mL cold (−20 °C) acetone. The protein pellet was recollected by centrifugation followed by two additional washing steps with 1 ml of acetone. Finally, the protein pellet was resuspended in 200 μL SDT buffer and subjected to a continued sonication treatment. The sample was centrifuged at 12,000 g for 15 min. The suspension was stored at −20 °C until use. Protein concentration was measured with BCA protein assay reagent. Subsequently, SDS-PAGE electrophoresis was performed to confirm the presence of protein bands (Fig. S1).

**In-solution digestion.** Protein digestion was performed according to the filter-aided sample preparation (FASP) procedure described by Wiśniewski *et al.*<sup>35</sup>. A protein sample (approximately 30 μg) was briefly solubilized in 30 μL of cell lysis buffer (4% SDS, 100 mM DTT, 150 mM Tris-HCl, pH 8.0) at 90 °C for 10 min. The detergent, DTT and other low-molecular-weight components were removed with 200 μL UA buffer (8 M Urea, 150 mM Tris-HCl, pH 8.0) via repeated ultrafiltration (Microcon units, 10 kD). Then 100 μL 50 mM iodoacetamide in UA buffer was added to block reduced cysteine residues. After 30 min incubation in darkness, the sample was collected by centrifugation at 14,000 g and room temperature for 20 min. The filter was washed with 100 μL UA buffer for three times and then with 100 μL 25 mM NH<sub>4</sub>HCO<sub>3</sub> for three times. Finally, the protein suspensions were digested with three enzymes (trypsin, GluC and chymotrypsin) in 40 μL 25 mM NH<sub>4</sub>HCO<sub>3</sub> at 37 °C overnight. The resulting peptides were collected as a filtrate.

**Shotgun LC-MS/MS analysis.** The shotgun LC-MS/MS were performed on a Q Exactive mass spectrometer coupled to Easy nLC (Proxeon Biosystems, now Thermo Fisher Scientific). Six μL of each fraction was injected for nanoLC-MS/MS analysis. The peptide mixture (5 μg) was loaded onto a C18-reversed phase column (Thermo Scientific Easy Column, 10 cm long, 75 μm inner diameter, 3 μm resin) in buffer A (0.1% formic acid) and separated with a linear gradient of buffer B (80% acetonitrile and 0.1% formic acid) at a flow rate of 250 nL/min controlled by IntelliFlow technology over 140 min. The MS data were acquired with a data-dependent top 10 method by dynamically choosing the most abundant precursor ions from the survey scan (300–1800 m/z) for HCD fragmentation. The determination of the target value was based on predictive automatic gain control. The dynamic exclusion duration was 60 s. The survey scans were acquired at a resolution of 70,000 at m/z 200, and the resolution for HCD spectra was set at 17,500 at m/z 200. The normalized collision energy was 29 eV and the underfill ratio, which specified the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.1%. The instrument was run by enabling the peptide recognition mode.

**Sequence Database Searching and Data Analysis.** MS/MS spectra were searched with MASCOT<sup>36</sup> engine (Matrix Science, London, UK, version 2.2) against the *E. octocarinatus* protein database including 32,353 protein entries. The following options were used to identify protein: peptide mass tolerance, 20 ppm; MS/MS

tolerance, 0.1 Da; enzyme = trypsin, chymotrypsin or GluC; missed cleavage, 2; fixed modification, carbamidomethyl (C); and variable modification, Oxidation (M). The filter parameters were protein FDR  $\leq$  0.01 and peptide FDR  $\leq$  0.01. Multiple peptide identifications were generally returned by SEQUEST for each MS/MS spectrum and for each parent-ion change state.

**Multiple sequence alignment.** DNASTar was used to perform multiple sequence alignment. Sequences with the following accession number were used to generate the alignment: NP\_004160.3 (*H. sapiens*), CCD63201.1 (*C. elegans*), and AAA21024.1 (*S. cerevisiae*).

## References

- Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip Rev RNA* **3**, 661–673 (2012).
- Drummond, D. A. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* **10**, 715–724 (2009).
- Caliskan, N., Peske, F. & Rodnina, M. V. Changed in translation: mRNA recoding by  $-1$  programmed ribosomal frameshifting. *Trends Biochem Sci* **40**, 265–274 (2015).
- Namy, O., Rousset, J. P., Napthine, S. & Brierley, I. Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* **13**, 157–168 (2004).
- Cobucci-Ponzano, B., Rossi, M. & Moracci, M. Translational recoding in archaea. *Extremophiles* **16**, 793–803 (2012).
- Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Recoding: translational bifurcations in gene expression. *Gene* **286**, 187–201 (2002).
- Chandler, M. & Fayet, O. Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol* **7**, 497–503 (1993).
- Brakier-Gingras, L., Charbonneau, J. & Butcher, S. E. Targeting frameshifting in the human immunodeficiency virus. *Expert Opin Ther Targets* **16**, 249–258 (2012).
- Jacobs, J. L., Belew, A. T., Rakauskaitė, R. & Dinman, J. D. Identification of functional, endogenous programmed  $-1$  ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **35**, 165–174 (2007).
- Belew, A. T. *et al.* Ribosomal frameshifting in the *CCR5* mRNA is regulated by miRNAs and the NMD pathway. *Nature* **512**, 265–269 (2014).
- Craigien, W. J. & Caskey, C. T. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* **322**, 273–275 (1986).
- Bekaert, M., Atkins, J. F. & Baranov, P. V. ARFA: a program for annotating bacterial release factor genes, including prediction of programmed ribosomal frameshifting. *Bioinformatics* **22**, 2463–2465 (2006).
- Ivanov, I. P. & Atkins, J. F. Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. *Nucleic Acids Res* **35**, 1842–1858 (2007).
- Kurian, L., Palanimurugan, R., Gödderz, D. & Dohmen, R. J. Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature* **477**, 490–U147 (2011).
- Zimmer, M., Sattelberger, E., Inman, R. B., Calendar, R. & Loessner, M. J. Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed  $+1$  translational frameshifting in structural protein synthesis. *Mol Microbiol* **50**, 303–317 (2003).
- Firth, A. *et al.* Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC\_UUU\_CGU and is in the  $+1$  direction. *Open Biol* **2**, 120109 (2012).
- Belcourt, M. F. & Farabaugh, P. J. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* **62**, 339–352 (1990).
- Farabaugh, P. J., Zhao, H. & Vimaiaadithan, A. A novel programmed frameshift expresses the *POL3* gene of retrotransposon Ty3 of Yeast: Frameshifting without tRNA slippage. *Cell* **74**, 93–103 (1993).
- Wang, R., Xiong, J., Wang, W., Miao, W. & Liang, A. High frequency of  $+1$  programmed ribosomal frameshifting in *Euplotes octocarinatus*. *Sci Rep* **6**, 21139 (2016).
- Klobutcher, L. A. Sequencing of random *Euplotes crassus* macronuclear genes supports a high frequency of  $+1$  translational frameshifting. *Eukaryot Cell* **4**, 2098–2105 (2005).
- Klobutcher, L. A. & Farabaugh, P. J. Shifty ciliates: frequent programmed translational frameshifting in Euplotids. *Cell* **111**, 763–766 (2002).
- Vallesi, A., Di Pretoro, B., Ballarini, P., Apone, F. & Luporini, P. A novel protein kinase from the ciliate *Euplotes raikovi* with close structural identity to the mammalian intestinal and male-germ cell kinases: Characterization and functional implications in the autocrine pheromone signaling loop. *Protist* **161**, 250–263 (2010).
- Candelori, A., Luporini, P., Alimenti, C. & Vallesi, A. Characterization and expression of the gene encoding *En-MAPK1*, an intestinal cell kinase (ICK)-like kinase activated by the autocrine pheromone-signaling loop in the polar ciliate, *Euplotes nobilii*. *Int J Mol Sci* **14**, 7457–7467 (2013).
- Aigner, S. *et al.* *Euplotes* telomerase contains an La motif protein produced by apparent translational frameshifting. *EMBO J* **19**, 6230–6239 (2000).
- Washburn, M. P., Wolters, D. & Yates, J. R. 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242–247 (2001).
- Tan, M., Heckmann, K. & Brunen-Nieweler, C. Analysis of micronuclear, macronuclear and cDNA sequences encoding the regulatory subunit of cAMP-dependent protein kinase of *Euplotes octocarinatus*: evidence for a ribosomal frameshift. *J Eukaryot Microbiol* **48**, 80–87 (2001).
- Karamysheva, Z. *et al.* Developmentally programmed gene elimination in *Euplotes crassus* facilitates a switch in the telomerase catalytic subunit. *Cell* **113**, 565–576 (2003).
- Jala, V. R., Appaji Rao, N. & Savithri, H. S. Identification of amino acid residues, essential for maintaining the tetrameric structure of sheep liver cytosolic serine hydroxymethyltransferase, by targeted mutagenesis. *Biochem J* **369**, 469–476 (2003).
- Meyer, F. *et al.* UGA Is Translated as Cysteine in pheromone-3 of *Euplotes octocarinatus*. *Proc Natl Acad Sci USA* **88**, 3758–3761 (1991).
- Turanov, A. A. *et al.* Genetic code supports targeted insertion of two amino acids by one codon. *Science* **323**, 259–261 (2009).
- Vallabhaneni, H., Fan-Minogue, H., Bedwell, D. M. & Farabaugh, P. J. Connection between stop codon reassignment and frequent use of shifty stop frameshifting. *RNA* **15**, 889–897 (2009).
- Fagan, C. E., Maehigashi, T., Dunkle, J. A., Miles, S. J. & Dunham, C. M. Structural insights into translational recoding by frameshift suppressor tRNA<sup>supf</sup>. *RNA* **20**, 1944–1954 (2014).
- Tuohy, T. M., Thompson, S., Gesteland, R. F. & Atkins, J. F. Seven, eight and nine-membered anticodon loop mutants of tRNA(2Arg) which cause  $+1$  frameshifting. Tolerance of DHU arm and other secondary mutations. *J Mol Biol* **228**, 1042–1054 (1992).
- Maehigashi, T., Dunkle, J. A., Miles, S. J. & Dunham, C. M. Structural insights into  $+1$  frameshifting promoted by expanded or modification-deficient anticodon stem loops. *Proc Natl Acad Sci USA* **111**, 12740–12745 (2014).

35. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359–362 (2009).
36. Perkins, D. N., Pappin, D. J., Creasy, D. M. & JS, C. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **18**, 3551–3567 (1999).
37. Tan, M., Liang, A. H., Brunen-Nieweler, C. & Heckmann, K. Programmed translational frameshifting is likely required for expressions of genes encoding putative nuclear protein kinases of the ciliate *Euplotes octocarinatus*. *J Eukaryot Microbiol* **48**, 575–582 (2001).
38. Mollenbeck, M., Gavin, M. C. & Klobutcher, L. A. Evolution of programmed ribosomal frameshifting in the *TERT* genes of *Euplotes*. *J Mol Evol* **58**, 701–711 (2004).
39. Wang, L. B., Dean, S. R. & Shippen, D. E. Oligomerization of the telomerase reverse transcriptase from *Euplotes crassus*. *Nucleic Acids Res* **30**, 4032–4039 (2002).
40. Doak, T. G., Witherspoon, D. J., Jahn, C. L. & Herrick, G. Selection on the genes of *Euplotes crassus* Tec1 and Tec2 transposons: Evolutionary appearance of a programmed frameshift in a Tec2 gene encoding a tyrosine family site-specific recombinase. *Eukaryot Cell* **2**, 95–102 (2003).

## Acknowledgements

This project is supported by grants from National Natural Science Foundation of China (No. 31372199) to A. Liang.

## Author Contributions

A.L. conceived and designed experiment; R.W. conducted the experiment. J.D. and Z.Z. guided data analysis; R.W. wrote the manuscript. A.L. and Y.F. edited the manuscript. All authors have read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Wang, R. *et al.* Large-scale mass spectrometry-based analysis of *Euplotes octocarinatus* supports the high frequency of +1 programmed ribosomal frameshift. *Sci. Rep.* **6**, 33020; doi: 10.1038/srep33020 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016