A Large Number of Nuclear Genes in the Human Parasite *Blastocystis* Require mRNA Polyadenylation to Create Functional Termination Codons

Vladimír Klimeš[1], Eleni Gentekaki[2], Andrew J. Roger[2,3], and Marek Eliáš[1,]*

[1]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Czech Republic

[2]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

[3]Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Halifax, Nova Scotia, Canada

*Corresponding author: E-mail: marek.elias@osu.cz.

Accepted: July 6, 2014

Abstract

Termination codons in mRNA molecules are typically specified directly by the sequence of the corresponding gene. However, in mitochondria of a few eukaryotic groups, some mRNAs contain the termination codon UAA deriving one or both adenosines from transcript polyadenylation. Here, we show that a similar phenomenon occurs for a substantial number of nuclear genes in *Blastocystis* spp., divergent unicellular eukaryote gut parasites. Our analyses of published genomic data from *Blastocystis* sp. subtype 7 revealed that polyadenylation-mediated creation of termination codons occurs in approximately 15% of all nuclear genes. As this phenomenon has not been noticed before, the procedure previously employed to annotate the *Blastocystis* nuclear genome sequence failed to correctly define the structure of the 3′-ends of hundreds of genes. From sequence data we have obtained from the distantly related *Blastocystis* sp. subtype 1 strain, we show that this phenomenon is widespread within the *Blastocystis* genus. Polyadenylation in *Blastocystis* appears to be directed by a conserved GU-rich element located four nucleotides downstream of the polyadenylation site. Thus, the highly precise positioning of the polyadenylation in *Blastocystis* has allowed reduction of the 3′-untranslated regions to the point that, in many genes, only one or two nucleotides of the termination codon are left.

Key words: *Blastocystis*, evolution, gene expression, mRNA processing, polyadenylation, termination codons, translation.

Introduction

The coding region of a mature mRNA molecule corresponds directly to the sequence of a primary transcript of a gene once intronic sequences are removed. However, some gene expression systems are more complex, and are most frequently reported within organelles of various eukaryotes. Unorthodox mechanisms for producing a functional coding sequence include posttranscriptional editing, found in mitochondrial, nuclear and plastid genes (Knoop 2011; Dorrell and Howe 2012; Takenaka et al. 2013), and *trans*-splicing recently discovered in some protist mitochondria (Moreira et al. 2012; Jackson and Waller 2013; Kiethega et al. 2013) and the nucleus of the diplomonad *Giardia intestinalis* (Kamikawa et al. 2011).

An interesting form of posttranscriptional sequence modification was described in mammalian mitochondria, in which part of the termination codon UAA of some mRNAs is created by polyadenylation (Anderson et al. 1981; Chang and Tong 2012); such mature transcripts therefore lack a 3′-untranslated region (3′-UTR). A similar phenomenon was also described for mRNAs of the *cox3* gene in the mitochondria of dinoflagellates (Jackson et al. 2007) and at least for some mRNAs encoded by the mitochondrial genome of the euglenozoan *Diplonema papillatum* (Kiethega et al. 2013). On the other hand, the occurrence of this phenomenon in nuclear transcripts seems to be very rare. To our knowledge, the only reported cases concern a few mammalian genes that employ alternative polyadenylation sites. One such case is the mouse thymidylate synthase gene that has multiple mRNA isoforms. The less abundant isoform exhibits the conventional gene-encoded termination codon UAG followed by a 3′-UTR and the poly(A) tail, whereas the more abundant isoform uses an alternative polyadenylation site that changes the UAG codon to UAA followed directly by the poly(A) tail (Jenh et al. 1986). Another case involves a mechanism dubbed polyadenylation-mediated Tyr-to-stop codon conversion (PAY*) that is used for several genes including glutamyl-prolyl tRNA synthetase. In this mechanism, an mRNA molecule is alternatively cleaved within an internal Tyr-encoding UAU or UAC codon, which is subsequently converted by poly(A) synthesis to the UAA termination codon, resulting in a C-terminally truncated protein variant (Yao et al. 2012). However, no eukaryotic nuclear gene that would lack a functional termination codon specified directly by the gene sequence has been found so far.

Blastocystis is a genus of anaerobic unicellular protists living in the intestine of various metazoans, including humans, as parasites or commensals (Clark et al. 2013). Indeed, it is probably the most prevalent protistan parasite in the human gut (Scanlan and Stensvold 2013). Its phylogenetic position had long been an enigma, but phylogenetic analyses of the 18S rRNA gene sequence (Silberman et al. 1996) and subsequent phylogenomic analyses of multigene matrices (e.g., Hampl et al. 2009) unambiguously demonstrated that it forms a lineage within a major eukaryotic assemblage called the Stramenopiles. *Blastocystis* cells have a number of unorthodox characteristics, the most notable of which are their mitochondrion-related organelles (MROs). Functionally, these organelles seem to lie halfway on a spectrum between canonical mitochondria of aerobic eukaryotes and anaerobic MROs called hydrogenosomes (Stechmann et al. 2008; Denoeud et al. 2011).

The unusual biological features of *Blastocystis* and its medical significance led to the initiation of a genome sequencing project, which was brought to fruition by Denoeud et al. (2011). Their draft genome sequence of *Blastocystis* sp. subtype 7 comprised around 18.8 Mb and harbored approximately 6,000 predicted protein-coding genes with short introns and intergenic regions. In this report, we describe a new striking aspect of *Blastocystis* molecular biology that we discovered while analyzing this genome.

Results and Discussion

We were initially interested in genes of the Ras superfamily of GTPases in the genome of *Blastocystis* sp. subtype 7. Many members of this superfamily undergo functionally essential posttranslational prenylation of one or two cysteine residues close to the C-terminus (Pechlivanis and Kuhlmann 2006). However, we noticed that some of the predicted *Blastocystis* Ras superfamily GTPases lack the expected C-terminal Cys-containing motifs (fig. 1*A* and supplementary fig. S1*A*, Supplementary Material online). By inspecting the expressed sequence tag (EST) sequences for these genes, we found that termination codons were created at the transcript level by the addition of poly(A) tails downstream of particular nucleotides in the primary transcripts. Once these termination codons are considered, the predicted proteins bear prenylation motifs typical for the corresponding group of GTPases (fig. 1*A* and supplementary fig. S1*A*, Supplementary Material online).

These observations prompted us to further investigate the extent of this phenomenon in this genome. Using available EST sequences from *Blastocystis* sp. subtype 7, we were able to define the polyadenylation site for 2,419 nuclear genes (additional 226 cases of potential polyadenylation were considered as spurious by a variety of criteria—see supplementary methods, Supplementary Material online). Of these genes, 372 (i.e., ~15%) unambiguously exhibited termination codons created by adding the poly(A) tail at a position where the underlying gene sequence did not specify a termination codon; in virtually all these genes, there was no evidence for an alternative form of the 3′-end of the transcript, suggesting that the introduction of the termination codons by polydenylation does not result from processing of aberrant or incomplete transcripts (e.g., by cytoplasmic adenylation involved in mRNA decay). Of the termination codons created this way, 359 were UAA and 13 were UGA. In addition, we identified 90 additional cases where the gene sequence specified potentially functional UAG or UGA termination codons, which were transformed into the UAA codon by polyadenylation (supplementary fig. S1*B*, Supplementary Material online).

Although the 372 genes with termination codons created by polyadenylation represent approximately 15% of all 2,419 nuclear genes with clearly defined polyadenylation sites, the number of poly(A)-containing ESTs associated with the former gene category was only 687, that is, approximately 4%, out of 16,006 poly(A)-containing ESTs mapping to the 2,419 nuclear genes. It is therefore possible that the genes with termination codons created by polyadenylation on average have a lower expression level than genes with termination codons specified directly by the gene sequence. We additionally tested the possibility that polyadenylation-mediated creation of termination codons occurs preferentially in certain functional gene categories. Using the Gene ontology (GO) framework (Ashburner et al. 2000), we found that the set of genes that use polyadenylation to create termination codons is significantly enriched (*P* value 2.0E-6, Fisher’s exact test) in genes annotated as GO:0003824—catalytic activity (i.e., genes encoding enzymes). The biological significance of this observation is unclear and analyses of all genes in the *Blastocystis* genome will be needed to confirm this bias.

Based on the foregoing estimate of approximately 15%, we assume that approximately 900 of the 6,000 annotated genes in this *Blastocystis* subtype rely on polyadenylation-mediated creation of termination codons. As this unusual phenomenon was not noticed by the annotation team of the *Blastocystis* sp. subtype 7 genome (Denoeud et al. 2011), the C-terminal regions of a large fraction of genes are predicted incorrectly; these erroneous gene models usually have either an artificial extension of the terminal exon to the first stop codon in-frame with the coding sequence (supplementary fig. S1*A*, Supplementary Material online) or an artificial intron accompanied by an artificial terminal exon that includes a hypothetical termination codon (e.g., see fig. 1*A*).

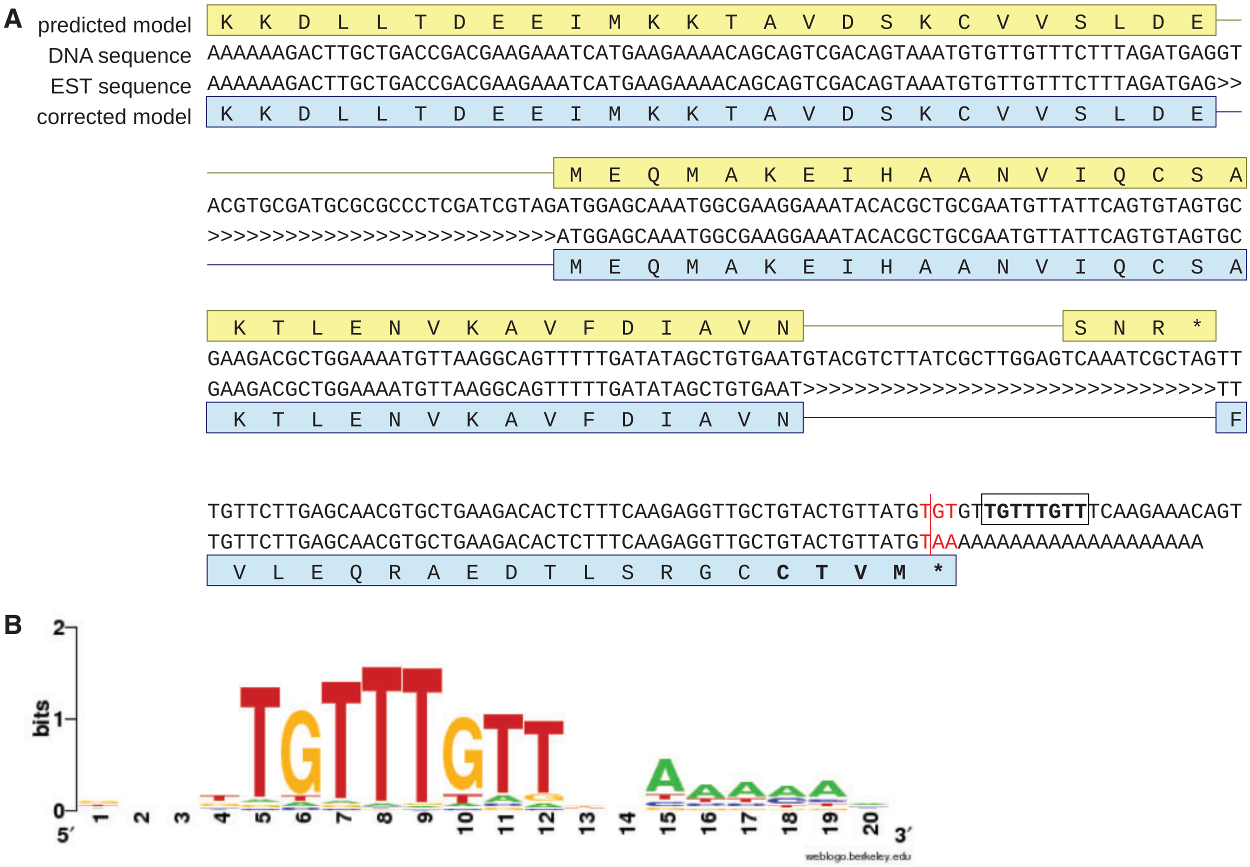
A
predicted model
DNA sequence
EST sequence
corrected model
K K D L L T D E E I M K K T A V D S K C V V S L D E
AAAAAAGACTTGCTGACCGACGAAGAAATCATGAAGAAAACAGCAGTCGACAGTAAATGTGTTGTTTCTTTAGATGAGGT
AAAAAAGACTTGCTGACCGACGAAGAAATCATGAAGAAAACAGCAGTCGACAGTAAATGTGTTGTTTCTTTAGATGAG>>
K K D L L T D E E I M K K T A V D S K C V V S L D E
M E Q M A K E I H A A N V I Q C S A
ACGTGCGATGCGCGCCCTCGATCGTAGATGGAGCAAATGGCGAAGGAAATACACGCTGCGAATGTTATTCAGTGTAGTGC
>>>>>>>>>>>>>>>>>>>>>>>>>>ATGGAGCAAATGGCGAAGGAAATACACGCTGCGAATGTTATTCAGTGTAGTGC
M E Q M A K E I H A A N V I Q C S A
K T L E N V K A V F D I A V N
S N R *
GAAGACGCTGGAAAATGTTAAGGCAGTTTTTGATATAGCTGTGAATGTACGTCTTATCGCTTGGAGTCAAATCGCTAGTT
GAAGACGCTGGAAAATGTTAAGGCAGTTTTTGATATAGCTGTGAAT>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>TT
K T L E N V K A V F D I A V N
F
TGTTCTTGAGCAACGTGCTGAAGACACTCTTTCAAGAGGTTGCTGTACTGTTATGTGTGTTGTTTGTTTCAAGAAACAGT
TGTTCTTGAGCAACGTGCTGAAGACACTCTTTCAAGAGGTTGCTGTACTGTTATGTAAAAAAAAAAAAAAAAAAAAA
V L E Q R A E D T L S R G C C T V M *
B
bits
2
1
0
5′
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
3′
weblogo.berkeley.edu

Fig. 1.—Polyadenylation-mediated creation of termination codons in *Blastocystis* sp. subtype 7. (*A*) An example of a gene from *Blastocystis* sp. subtype 7 with polyadenylation-dependent creation of a termination codon. The figure shows the 3′-end of a gene (FN668683.1, positions 539586–539266) encoding a GTPase of the RHO family, which is characterized by the presence of a C-terminal CXXX motif. Although the current prediction of the respective protein sequence (CBK24281.2) lacks the expected C-terminal motif, the protein sequence predicted by taking into account an EST sequence (FQ805196.1) does exhibit the motif (CTVM in bold). The termination codon in the EST sequence (TAA) created by polyadenylation is highlighted in red, the conserved polyadenylation motif (see the text) is boxed, introns are marked by the characters ">." The sequences displayed in the figure are defined above by their GenBank accession numbers. (*B*) A sequence logo for the conserved motif downstream from the polyadenylation site. The logo was created using WebLogo (Crooks et al. 2004; http://weblogo.berkeley.edu/, last accessed July 15, 2014) on the basis of 2,419 individual sequences.

We then looked for possible signals that may govern the positioning of the polyadenylation sites in *Blastocystis*. Studies of the mRNA 3′-end processing in metazoan cells established the existence of a highly conserved AAUAAA motif located 20–30 nucleotides upstream of the polyadenylation site and a less well-conserved GU-rich sequence (also called downstream sequence element, DSE) (Proudfoot 2011). Although we could not find any such conserved motif upstream of the polyadenylation sites in *Blastocystis*, we identified the conserved motif TGTTTGTT (UGUUUGUU at the RNA level), or close variants (fig. 1*B*), located four (nonconserved) nucleotides downstream of the polyadenylation site (fig. 1*A* and supplementary fig. S1, Supplementary Material online). This sequence pattern was highly conserved: Of the 2,419 nuclear genes with clearly defined polyadenylation sites, 2,333 (i.e., ~96%) exhibited the motif located exactly four nucleotides downstream the polyadenylation site and differing from the TGTTTGTT consensus by up to two substitutions, whereas the remaining approximately 4% of cases displayed three or four substitutions in the motif and/or the motif was located three of five (instead of four) nucleotides downstream of the polyadenylation site. The conservation of the motif and its invariant distance from the polyadenylation site enabled us to revise the predictions of the 3′-end of some *Blastocystis* genes currently lacking direct transcriptomic evidence (examples are shown in fig. 2 and supplementary fig. S1C and *D*, Supplementary Material online).

We hypothesize that the TGTTTGTT motif functions similarly to DSE in mammals, which recruits a multisubunit complex, the cleavage stimulation factor (CstF), through the binding of its subunit CSTF2 (also called CstF-64) (Pérez Cañadillas and Varani 2003). To test this notion we looked

```
subtype 7       DNA AACGCGTCTACGCAGTTCGAGTGTGGATATGCGCGAAGACAAACTTTTACCTCCCAATCCAATGGCTTTAATTATATCTTTTTTAATTGTTGTTATTGTTATTGTGTTTTTTGAGTGTTT
    predicted model  T  R  L  R  S  S  S  V  D  M  R  E  D  K  L  L  P  P  N  P  M  A  L  I  I  S  F  L  I  V  V  I  V  I  V  F  F  E  C  F
P. infestans         A  R  R  R  S  S  S  V  D  L  S  D  N  D  P  T  -  P  S  A  I  H  I  L
              model  S  R  V  R  S  S  S  I  D  I  R  D  E  R  V  I  P  P  N  P  M  S  I  I
subtype 1       EST CTCGCGCGTGCGCAGCTCCAGCATCGACATCCGAGACGAGCGGGTCATTCCACCCAATCCCATGTCCATCATC>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>-------
                DNA CTCGCGCGTGCGCAGCTCCAGCATCGACATCCGAGACGAGCGGGTCATTCCACCCAATCCCATGTCCATCATCATATCCTTGTGCCCTGCGACATTTTATTTCTTGACGGCAG-------

                    CTTTAACTTGGTGTACGGAGATATTAGCGTAGATGGACTTGGCTTGATCAAGTGGTGGATTCCCGATGAATCCGAGCAATTTTGTGTTTGTTGCAAACAGAACGAACGCTAAATCGACTA
                     F  N  L  V  Y  G  D  I  S  V  D  G  L  G  L  I  K  W  W  I  P  D  E  S  E  Q  F  C  V  C  C  K  Q  N  E  R  *
                                    A  T  I  D  E  D  G  L  G  L  S  S  W  W  T  S  *
                                    G  D  I  S  Y  D  G  L  G  L  I  K  W  W  I  P  E  E  *
                    ----------------GGAGATATAAGCTACGATGGACTTGGCTTGATCAAGTGGTGGATTCCCGCGGAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCT
                    ----------------GGAGATATAAGCTACGATGGACTTGGCTTGATCAAGTGGTGGATTCCCGAGGAATCGAGTGTTTGTTGCAAACAGCGATCTAAAAGGACTACTGCTTTCCCGAC
```

Fig. 2.—Example of a gene from *Blastocystis* sp. subtype 1 displaying polyadenylation-mediated creation of the termination codon. The termination codon in the transcript (TAA) is highlighted in red, the putative polyadenylation motif is boxed, an intron is marked by the characters ">." Homologous proteins from *Blastocystis* sp. subtype 7 and *Phytophthora infestans* are shown for comparison. Note that the protein sequence from *Blastocystis* sp. subtype 7 is most likely incorrectly predicted, because the gene also seems to rely on polyadenylation-mediated creation of a termination codon, as is indicated by the presence of a putative polyadenylation motif (boxed). Hence, the triplet TTT (in red) is probably changed to TAA in a corresponding transcript, which would shorten the predicted coding sequence of the gene and make the encoded protein more similar to its homologs. Subtype 7 DNA sequence: FN668683.1, positions 458734–458973; subtype 7 predicted model: CBK24259.2; *Phytophthora infestans* protein sequence: XP_002895556.1; subtype 1 DNA sequence: a part of a scaffold from a genome assembly based on 454 and Illumina reads (Gentekaki E, Curtis B, Roger AJ, unpublished data); subtype 1 EST sequence: EC650050.1; subtype 1 predicted protein sequence: a theoretical inference based on the EST sequence.

for a possible ortholog of the human CSTF2 in *Blastocystis* and found a protein, CBK24645.2, that is the reciprocal best BLAST hit to CSTF2. However, the *Blastocystis* protein is much shorter (142 amino acid residues) than CSTF2 (577 amino acid residues) and corresponds only to its N-terminal region comprising the RNA-binding domain and the so-called hinge domain (Hockert et al. 2010). No EST evidence is available to confirm that the gene model underlying the putative CSTF2 ortholog in *Blastocystis* sp. subtype 7 is correct, but investigation of the genomic sequence downstream of the gene did not reveal a region potentially encoding the missing C-terminal part of the protein. Our unpublished transcriptomic data from *Blastocystis* sp. subtype 1 indicate that CSTF2 is indeed a short protein confined to the two N-terminal domains in the *Blastocystis* genus (data not shown). The presence of the RNA-binding domain suggests that this protein may potentially be the binding partner of the TGTTT GTT motif, but the lack of the C-terminal region including a functionally important C-terminal domain conserved at least in opisthokonts and plants (Qu et al. 2007) indicates that there may be important differences in mechanistic details of the CstF functioning between *Blastocystis* and other eukaryotes. A broader comparative analysis and direct experimental characterization of the mRNA 3′-end processing machinery in *Blastocystis* are needed to allow more specific conclusions.

As *Blastocystis* is a genetically diverse genus (Clark et al. 2013), it is of interest to know whether polyadenylation-mediated termination codon creation in nuclear mRNA transcripts is specific to subtype 7 or whether it is more widespread. To address this, we used previously published EST data (Stechmann et al. 2008) to investigate our preliminary draft genome sequence of *Blastocystis* sp. subtype 1 for the presence of genes that require this mechanism. Indeed, we found a number of cases of poly(A)-mediated termination codon creation, an example of which is portrayed in figure 2. In this case, a triplet TCG in the sequence of a gene encoding a homolog of the RPC8 subunit of the RNA polymerase III is changed in a corresponding transcript to an in-frame termination codon TAA. Furthermore, the position of the polyadenylation site in *Blastocystis* sp. subtype 1 is apparently governed by the same conserved polyadenylation motif as in the subtype 7 (fig. 2). Therefore, it is likely that poly(A)-mediated generation of termination codons is a feature that evolved either prior to, or early within, the diversification of the major subtypes of the *Blastocystis* genus. It will be interesting to investigate whether a similar mechanism occurs in related anaerobic stramenopile groups, such as opalinids and proteromonads (Silberman et al. 1996; Kostka et al. 2007).

Conclusions

Termination codons are essential signals in mRNA molecules and their absence triggers a specific mechanism of mRNA surveillance called nonstop decay (Klauer and van Hoof 2012). It appears that in the *Blastocystis* lineage the specification of the polyadenylation site has evolved to exhibit at least three unusual properties: 1) It does not utilize a conserved upstream element; 2) it relies on a highly conserved DSE, whereas other eukaryotes in which the mRNA 3′-end processing mechanism has been investigated in sufficient detail exhibit a much higher level of degeneracy of DSE (Wahle and Rüegsegger 1999; Proudfoot 2011); and 3) it is so precise (i.e., four nucleotides upstream of the highly conserved DSE) that over time the 3′-UTRs in many genes were reduced until all that was left was the first one or two nucleotides of the termination codon. Comprehensive genomic and transcriptomic investigations of diverse *Blastocystis* lineages are needed to better understand the evolutionary origin of this phenomenon, and biochemical studies are required to elucidate the mechanistic details of the 3′-end processing of primary transcripts in *Blastocystis*.

Materials and Methods

Briefly, the analysis used the *Blastocystis* sp. subtype 7 genome assembly version 2.0 (https://www.genoscope.cns.fr/externe/Download/Projets/Blastocystis/assembly/, last accessed July 15, 2014) and EST sequences (34,470 in total) submitted to GenBank (http://www.ncbi.nlm.nih.gov, last accessed July 15, 2014) by Genoscope. The program STAR (Dobin et al. 2013) was used to align the EST sequences to the genome sequence. In-house programs written in the Java language were then used to define polyadenylation sites and the cases where polyadenylation introduces a termination codon into the transcript. Genes exhibiting polyadenylation-mediated creation of termination codons in the *Blastocystis* sp. subtype 1 were identified in a draft genome assembly obtained from 454 and Illumina reads (Gentekaki E, Curtis B, Roger AJ, unpublished data) using EST data available for this isolate (Stechmann et al. 2008). The GO term enrichment analysis was performed using the Blast2GO tool (Conesa et al. 2005). Conserved motifs that may constitute the polyadenylation signal in a 50-bp region upstream and downstream of the mapped polyadenylation sites were searched using the qPMS7 algorithm for de novo motif search (Dinh et al. 2012; http://motifsearch.com/, last accessed July 15, 2014). Details on the analyses are provided in supplementary methods, Supplementary Material online.

Supplementary Material

Supplementary methods and figure S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Patrick J. Keeling, Matúš Valach, and several anonymous reviewers for valuable comments that helped to improve this article. This work was supported by the Czech Science Foundation Grant 13-24983S, Structural Funds of the EU (Project CZ.1.05/2.1.00/03.0100 IET) and the Project LO1208 of the National Feasibility Programme I of the Czech Republic to M.E. and the Canadian Institutes of Health Research Operating Grant MOP-62809 to A.J.R.

Literature Cited

Anderson S, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Ashburner M, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25:25–29.

Chang JH, Tong L. 2012. Mitochondrial poly(A) polymerase and polyadenylation. Biochim Biophys Acta. 1819:992–997.

Clark CG, van der Giezen M, Alfellani MA, Stensvold CR. 2013. Recent developments in *Blastocystis* research. Adv Parasitol. 82:1–32.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

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

Denoeud F, et al. 2011. Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. Genome Biol. 12:R29.

Dinh H, Rajasekaran S, Davila J. 2012. qPMS7: a fast algorithm for finding (ℓ, d)-motifs in DNA and protein sequences. PLoS One 7:e41425.

Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Dorrell RG, Howe CJ. 2012. Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. Proc Natl Acad Sci U S A. 109:18879–18884.

Hampl V, et al. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc Natl Acad Sci U S A. 106:3859–3864.

Hockert JA, Yeh HJ, MacDonald CC. 2010. The hinge domain of the cleavage stimulation factor protein CstF-64 is essential for CstF-77 interaction, nuclear localization, and polyadenylation. J Biol Chem. 285:695–704.

Jackson CJ, Waller RF. 2013. A widespread and unusual RNA trans-splicing type in dinoflagellate mitochondria. PLoS One 8:e56777.

Jackson CJ, et al. 2007. Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. BMC Biol. 5: 41.

Jenh CH, Deng TL, Li DW, DeWille J, Johnson LF. 1986. Mouse thymidylate synthase messenger RNA lacks a 3′ untranslated region. Proc Natl Acad Sci U S A. 83:8482–8486.

Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, Hashimoto T. 2011. Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated trans-splicing. Curr Biol. 21:311–315.

Kiethega GN, Yan Y, Turcotte M, Burger G. 2013. RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. RNA Biol. 10: 301–313.

Klauer AA, van Hoof A. 2012. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. Wiley Interdiscip Rev RNA. 3: 649–660.

Knoop V. 2011. When you can't trust the DNA: RNA editing changes transcript sequences. Cell Mol Life Sci. 68:567–586.

Kostka M, Cepicka I, Hampl V, Flegr J. 2007. Phylogenetic position of *Karotomorpha* and paraphyly of Proteromonadidae. Mol Phylogenet Evol. 43:1167–1170.

Moreira S, Breton S, Burger G. 2012. Unscrambling genetic information at the RNA level. Wiley Interdiscip Rev RNA. 3:213–228.

Pechlivanis M, Kuhlmann J. 2006. Hydrophobic modifications of Ras proteins by isoprenoid groups and fatty acids—more than just membrane anchoring. Biochim Biophys Acta. 1764:1914–1931.

Pérez Cañadillas JM, Varani G. 2003. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. EMBO J. 22: 2821–2830.

Proudfoot NJ. 2011. Ending the message: poly(A) signals then and now. Genes Dev. 25:1770–1782.

Qu X, et al. 2007. The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3′-end processing. J Biol Chem. 282:2101–2115.

Scanlan PD, Stensvold CR. 2013. *Blastocystis*: getting to grips with our guileful guest. Trends Parasitol. 29:523–529.

Silberman JD, Sogin ML, Leipe DD, Clark CG. 1996. Human parasite finds taxonomic home. Nature 380:398.

Stechmann A, et al. 2008. Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr Biol. 18: 580–585.

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. 2013. RNA editing in plants and its evolution. Annu Rev Genet. 47:335–352.

Wahle E, Rüegsegger U. 1999. 3′-End processing of pre-mRNA in eukaryotes. FEMS Microbiol Rev. 23:277–295.

Yao P, et al. 2012. Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. Cell 149: 88–100.

Associate editor: Geoff McFadden