Received: 27 March 2018 | Revised: 26 April 2018 | Accepted: 30 April 2018

DOI: 10.1002/wrna.1487

WILEY WIREs RNA

ADVANCED REVIEW

High throughput sequencing revolution reveals conserved fundamentals of U-indel editing

Sara L. Zimmer[1] | Rachel M. Simpson[2] | Laurie K. Read[2]

[1]Department of Biomedical Sciences, University of Minnesota Medical School, Duluth, Minnesota

[2]Department of Microbiology and Immunology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York

Correspondence
Dr Sara L. Zimmer, Department of Biomedical Sciences, University of Minnesota Medical School, Duluth, MN.
Email: szimmer3@d.umn.edu

Funding information
National Institutes of Health, Grant/Award Number: AI061580; American Heart Association, Grant/Award Number: 16SDG26420019

Among Euglenozoans, mitochondrial RNA editing occurs in the diplonemids and in the kinetoplastids that include parasitic trypanosomes. Yet U-indel editing, in which open reading frames (ORFs) on mRNAs are generated by insertion and deletion of uridylates in locations dictated by guide RNAs, appears confined to kinetoplastids. The nature of guide RNA and edited mRNA populations has been cursorily explored in a surprisingly extensive number of species over the years, although complete sets of fully edited mRNAs for most kinetoplast genomes are largely missing. Now, however, high throughput sequencing technologies have had an enormous impact on what we know and will learn about the mechanisms, benefits, and final edited products of U-indel editing. Tools including PARERS, TREAT, and *T-Aligner* function to organize and make sense of U-indel mRNA transcriptomes, which are comprised of mRNAs harboring uridylate indels both consistent and inconsistent with translatable products. From high throughput sequencing data come arguments that partially edited mRNAs containing "junction regions" of noncanonical editing are editing intermediates, and conversely, arguments that they are dead-end products. These data have also revealed that the percent of a given transcript population that is fully or partially edited varies dramatically between transcripts and organisms. Outstanding questions that are being addressed include the prevalence of sequences that apparently encode alternative ORFs, diversity of editing events in ORF termini and 5′ and 3′ untranslated regions, and the differences that exist in this byzantine process between species. High throughput sequencing technologies will also undoubtedly be harnessed to probe U-indel editing's evolutionary origins.

This article is categorized under:
RNA Processing > RNA Editing and Modification
RNA Evolution and Genomics > Computational Analyses of RNA

KEYWORDS
constructive neutral evolution, evolvability, parasitology, RNA editing core complex, RNA editing substrate binding complex

1 | INTRODUCTION

A radical and unique type of RNA editing occurs in the mitochondria of the kinetoplastid protozoa (Aphasizheva & Aphasizhev, 2016; Hashimi, Zimmer, Ammerman, Read, & Lukeš, 2013; Read, Lukeš, & Hashimi, 2016). Uridylates (U) are inserted or deleted in targeted locations in mRNAs in a process called U-indel editing, using largely *trans*-acting mitochondrially encoded template RNAs called guide RNAs (gRNAs) to direct the location and nature of the modification (Figure 1a). The

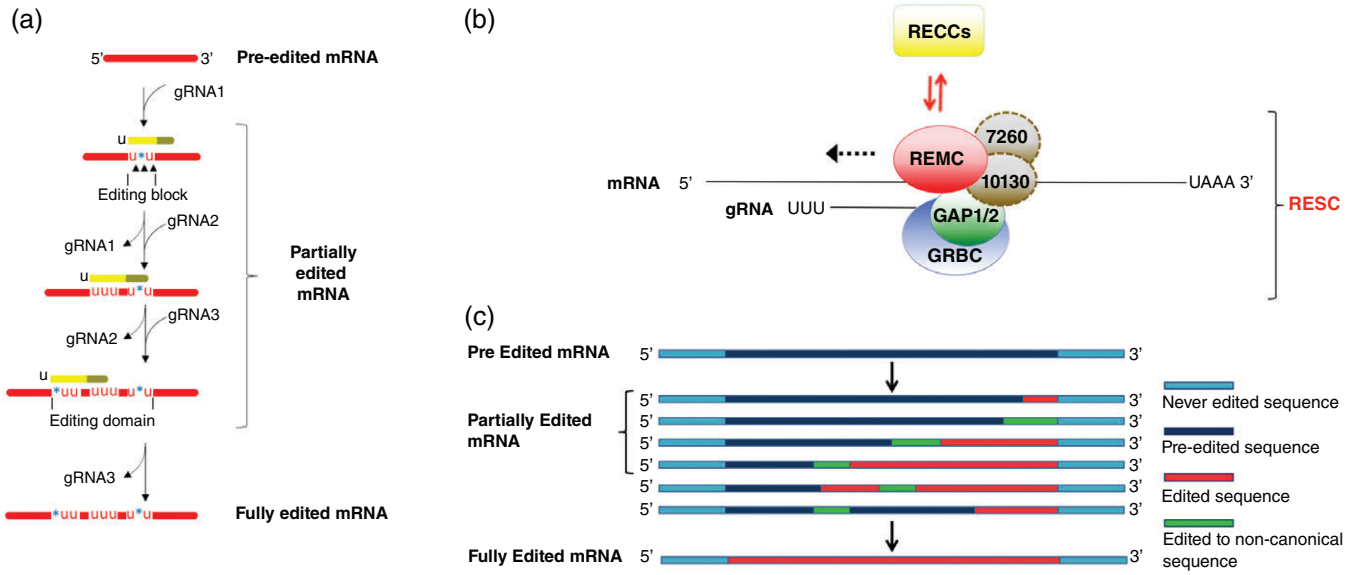

FIGURE 1 Features of U-indel editing. (a) U-indel editing is a process of insertion and deletion events guided by gRNAs that contain both regions that anneal to their cognate mRNAs, and regions that will direct editing of adjacent sequence. (Reprinted with permission from Read et al. (2016). Copyright 2016 Wiley Periodicals, Inc.) (b) Schematic diagram depicting the associations of the RNA editing mediator complex (REMC) with the guide RNA binding complex (GRBC), and other proteins to form the RNA editing substrate binding complex (RESC). Once these protein players have assembled with mRNA and the necessary gRNA, interaction with the catalytic components of editing, the RNA edited core complexes (RECCs), can proceed. GAP1/2 are gRNA binding proteins within GRBC, and the brown dotted-outline circles represent RESC proteins MRB7260 and MRB10130 that are not strictly REMC or GRBC components (McAdams, Simpson, Chen, Sun, & Read, 2018; Read et al., 2016). (c) Depiction of different types of mRNA molecules with origins in a single pan-edited cryptogene. Extreme 5′ and 3′ regions of the transcript never get edited. The 3′ never edited region is the binding site of the first gRNA of the pan-editing process. Pre-edited sequence exists in regions yet to be edited in order to generate an mRNA encoding a canonical protein product. The bracketed forms of the mRNA represent states that are consistent with editing in progress. In these examples, the green color exists at the boundary of the editing process (junction sequence) and the molecule may represent an editing intermediate that will continue to undergo editing to eventually achieve the fully edited state shown at the bottom. The unbracketed molecules are examples of regions of alternative editing that are not in a junction position. The alternatively edited region within the canonically edited region would result in product in which the edited region contains noncanonical codons and/or a frameshift, or else a misedited product. The region of noncanonical editing upstream of the main progression of editing represents a product that is likely misedited

insertion and deletion of U results in the repair of frameshifts, creation of stop and start codons, and often the generation of complete open reading frames (ORFs) compared to the "cryptogenes" in the mitochondrial genome. Thus, the RNA editing that takes place in the kinetoplastid mitochondrion is far more drastic than the modification of one nucleotide that defines most other types of mRNA editing (Licht & Jantsch, 2016; Nisbet, Kurniawan, Bowers, & Howe, 2016; Takenaka, Zehrmann, Verbitskiy, Härtel, & Brennicke, 2013; Valach, Moreira, Hoffmann, Stadler, & Burger, 2017; Yablonovitch, Deng, Jacobson, & Li, 2017).

Among Euglenozoan major lineages, mitochondrial RNA U modification is absent among euglenids but observed in the diplonemids and kinetoplastids (Faktorova, Dobakova, Pena-Diaz, & Lukeš, 2016). Basic knowledge of U-appendage RNA editing (causing insertion only) between spliced fragments of mitochondrial mRNAs and rRNAs in diplonemids is only now beginning to emerge (Valach et al., 2017). In contrast, an intense focus on U-indel editing in kinetoplastids has already led to admirable progress in elucidating the mechanism whereby it occurs, and the enzymes and protein factors involved. Briefly, U-indel editing requires an extensive collection of molecular complexes, including three related catalytic RNA editing core complexes (RECCs; a.k.a., 20S editosomes) that transiently interact with an incompletely defined RNA editing substrate binding complex (RESC; a.k.a. MRB) (Figure 1b). Several recent U-indel editing reviews (Aphasizheva & Aphasizhev, 2016; Hashimi et al., 2013; Read et al., 2016) focus on these protein complexes responsible for U-indel editing and rely almost exclusively on the model kinetoplastid, *Trypanosoma brucei.* Here, we examine the nature of U-indel edited transcriptomes (i.e., the edited sequences themselves) in a wide range of kinetoplastids, with a focus on the enormous impact and promise of high throughput sequencing technologies now being applied to this byzantine process. Application of these technologies is well underway, as demonstrated by the development of three bioinformatics tools dedicated to U-indel genome analysis, developed by independent teams (Carnes et al., 2017; Gerasimov et al., 2018; Simpson, Bruno, Bard, Buck, & Read, 2016). Indeed, these and other high throughput sequencing approaches (Aphasizheva et al., 2014; Kirby, Sun, Judah, Nowak, & Koslowsky, 2016; Koslowsky, Sun, Hindenach, Theisen, & Lucas, 2014; Madina et al., 2014) have transformed our ability to address U-indel editing's most fascinating paradoxes, including its very existence.

2 | EARLY STUDIES MAP THE BASIC U-INDEL EDITED TRANSCRIPT LANDSCAPE

The kinetoplastid mitochondrial transcriptome is exceedingly complex and contains RNAs transcribed from two types of mitochondrial circular DNA molecules, the maxicircles and the minicircles (Jensen & Englund, 2012; Povelones, 2014). Dozens of nearly identical maxicircles encode the mRNAs, some or all of which (depending on the organism) require U-indel editing to generate ORFs. Maxicircles can encode three classes of mRNAs: (a) never edited mRNAs that are correctly encoded in the genome, (b) minimally edited mRNAs requiring editing within small domains, and (c) pan-edited mRNAs requiring editing throughout their lengths and for which editing generates the entire ORF. The vast majority of gRNAs are encoded on the minicircles that cluster into hundreds of sequence classes (Kirby et al., 2016; Koslowsky et al., 2014). Minicircles also encode small RNAs that appear to play a role in gRNA 3′ end processing (Suematsu et al., 2016) and await further characterization. To this day, complete sets of fully edited mRNAs for kinetoplastid genomes are largely missing (Figure 2), and uncertainties abound even in well-studied genomes (Read, Wilson, et al., 1994). Finally, very few studies have targeted gRNA or small RNA populations (Aphasizheva et al., 2014; Kirby et al., 2016; Koslowsky et al., 2014; Madina et al., 2014).

The edited sequences of maxicircle-encoded mRNAs were initially determined by studies in the late 1980s and early 1990s, mainly in *T. brucei* and *Leishmania tarentolae*, which employed Sanger sequencing of multiple cDNA libraries. In some cases, findings were confirmed or extended by direct RNA sequencing (Figure 2 and references therein). To define consensus edited sequences of pan-edited mRNAs, cDNA libraries were initially interrogated by 3′ rapid amplification of cDNA ends (RACE), using a 5′ primer corresponding to a pre-edited mRNA sequence near the predicted 5′ ends of a given gene. Alignment of dozens to nearly 100 cDNA clones per gene revealed a conserved feature of the editing process. Namely, the ubiquitous recovery of mRNAs with unedited 5′ ends and 3′ ends containing progressively edited sequence in all studies demonstrated that editing proceeds in a general 3′ to 5′ direction on the mRNA (Figure 1). The consensus edited sequence of the 3′ end was identified from these partially edited cDNA clones. These 3′ consensus edited regions within cDNAs were usually bounded at their 5′ ends by variable nonconsensus edited sequences of differing lengths that were termed junctions (Koslowsky, Bhat, Read, & Stuart, 1991; Figure 1a,c). Junction regions are 5′ bounded by pre-edited sequence. Consensus edited sequence present at the 3′ ends of a majority of cDNA clones was then used to construct a 3′ primer for 5′ RACE and generation of a second cDNA library. This process was repeated iteratively until a 5′ edited consensus sequence was determined. Early studies also revealed that edited mRNAs typically contain a poly(A) tail with sporadic and nonconsensus U insertions, and that pan-edited mRNAs possess short never edited regions on their extreme ends. The lengths of never edited regions range from 25 to 40 nt on the 5′ end and are approximately 15 nt on the 3′ end. The function of the 3′ never edited region was revealed upon the identification of gRNAs: it is the region to which the first gRNA anchors; the presence of the 5′ never edited region remains mysterious.

Conventional analyses of edited mRNAs and gRNAs also hinted at the mechanisms and regulation of editing. For example, northern blot analyses demonstrated developmentally regulated accumulation of some edited mRNAs between insect procyclic and mammalian bloodstream stage *T. brucei* (Corell et al., 1994; Feagin et al., 1987; Feagin & Stuart, 1988; Koslowsky et al., 1990; Read et al., 1992; Souza et al., 1992) and *Trypanosoma congolense* (Read, Stankey, Fish, Muthiani, & Stuart, 1994). For *CR*4 and *ND7* mRNAs, differential editing appeared to be controlled at the level of 3′ to 5′ editing progression, as full length consensus sequences could be determined only in bloodstream stages, whereas 3′ partially edited cDNAs were acquired from both stages (Corell et al., 1994; Koslowsky et al., 1990). Similarly, northern blotting of specific gRNAs revealed increased abundances in life cycle stages in which edited mRNAs were decreased, suggesting that editing leads to gRNA destruction (Koslowsky, Riley, Feagin, & Stuart, 1992; Riley, Corell, & Stuart, 1994). Finally, sequencing of cDNA libraries revealed alternatively edited mRNA populations that could lead to production of more than one protein, as discussed in greater depth in the following section (Ochsenreiter & Hajduk, 2006; Read et al., 1992; Read, Wilson, et al., 1994). Despite these findings, practical collection limits in the numbers of cDNA and gRNA sequences confined the strength of these conclusions. In addition to the extensive *T. brucei* and *L. tarentolae* investigations, U-indel editing of mitochondrial transcripts has been analyzed to varying extents in an astounding 34 different kinetoplastid species (Figure 2). For a majority of these, the study's target was one or a very few gene products, and sometimes editing was inferred by comparison of the DNA sequence of the species in question with the edited RNA from another species. High throughput sequencing promises to rapidly and significantly expand our understanding of the process of U-indel editing as well as the breadth and variations in U-indel editing between wide-ranging kinetoplastids.

3 | HARNESSING HIGH THROUGHPUT SEQUENCING

High throughput sequencing technologies have exponentially increased sequence numbers obtainable in a single experiment, thus allowing the study of the whole mitochondrial transcriptome and the investigation of previously unanswerable questions

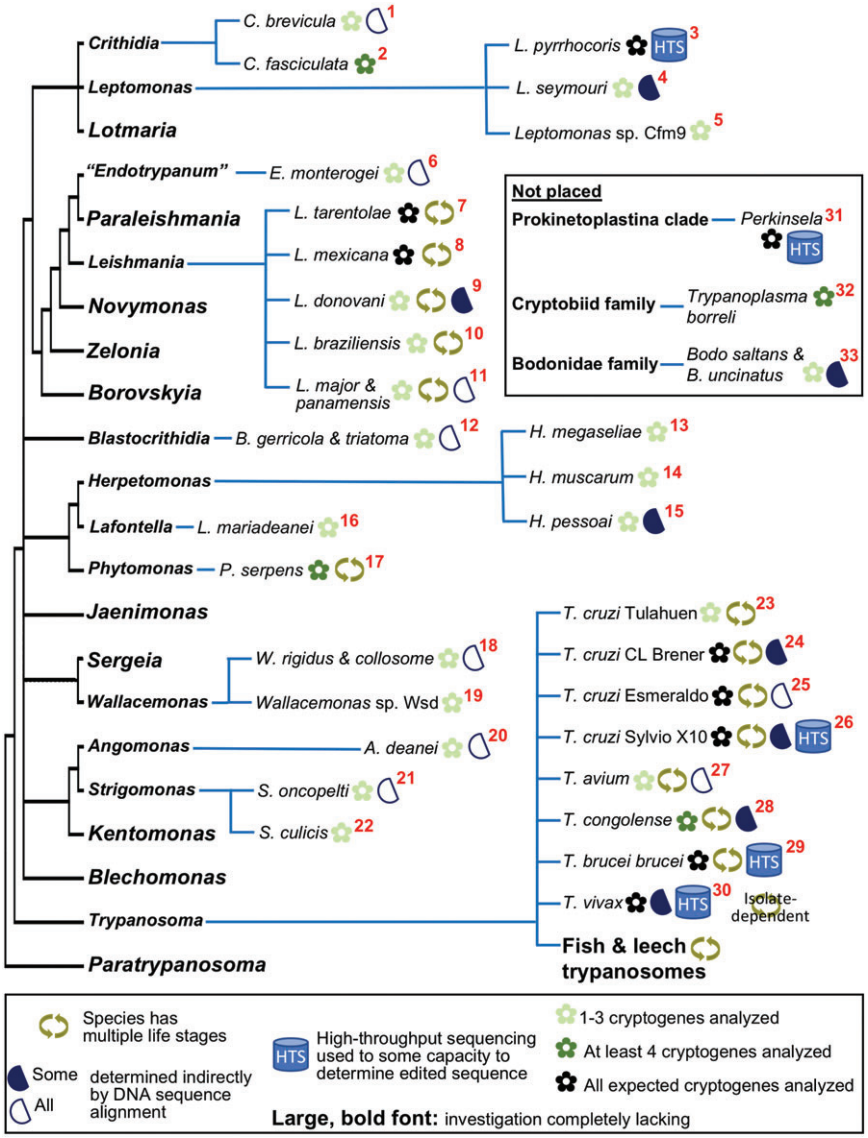

FIGURE 2 Current state of investigation of U-indel edited sequences from maxicircle cryptogenes among the kinetoplastids. Included studies are those in which RNA and/or DNA sequence is provided in text, figures, or supplementary material, that describe or compare pre-, partially, or fully edited sequences of maxicircle cryptogenes. Unless indicated with the half-moon symbol, RNA has been sequenced directly, cDNA has been cloned and sequenced, or amplified PCR products from cDNA templates have been cloned and sequenced. The structure of the tree is approximately based on models presented in Lukeš et al. (2018). Numbers denote the reference(s) for the analysis/es: 1. Merzlyak, Zakharova, and Kolesnikov (2001); 2. Benne et al. (1986), Feagin, Shaw, Simpson, and Stuart (1988), Maslov, Avila, Lake, and Simpson (1994), Shaw, Feagin, Stuart, and Simpson (1988), van der Spek et al. (1988), Yasuhira and Simpson (1995); 3. Gerasimov et al. (2018); 4. Bessolitsyna, Fediakov, Merzliak, and Kolesnikov (2005), Kolesnikov, Merzliak, Bessolitsyna, Fediakov, and Shoenian (2003), Merzlyak et al. (2001); 5. Kolesnikov et al. (2003); 6. Blom et al. (1998); 7. Bhat, Myler, and Stuart (1991), Shaw et al. (1988), Maslov et al. (1992), Shaw, Campbell, and Simpson (1989), Shaw et al. (1988), Souza, Myler, and Stuart (1992), Thiemann, Maslov, and Simpson (1994); 8. Maslov (2010); 9. Kolesnikov et al. (2003), Nebohácová, Kim, Simpson, and Maslov (2009); 10. Ramírez, Puerta, and Requena (2011); 11. Kolesnikov et al. (2003); 12. Merzlyak et al. (2001); 13. Landweber, Fiks, and Gilbert (1993), Landweber and Gilbert (1993), Maslov et al. (1994); 14. Landweber and Gilbert (1993), Maslov et al. (1994); 15. Kolesnikov et al. (2003), Landweber and Gilbert (1993), Merzlyak et al. (2001); 16. Landweber and Gilbert (1993); 17. Maslov, Hollar, Haghighat, and Nawathean (1998), Maslov, Nawathean, and Scheel (1999), Nawathean and Maslov (2000); 18. Gerasimov, Kostygov, Yan, and Kolesnikov (2012), Kolesnikov et al. (2003); 19. Gerasimov et al. (2012); 20. Gerasimov et al. (2012); 21. Aravin, Yurchenko, Merzlyak, and Kolesnikov (1998), Kolesnikov et al. (2003); 22. Maslov et al. (1994); 23. Kim, Teixeira, Kirchhoff, and Donelson (1994), Ochs, Otsu, Teixeira, Moser, and Kirchhoff (1996); 24. Ruvalcaba-Trejo and Sturm (2011), Shaw, Kalem, and Zimmer (2016), Thomas, Martinez, Westenberger, and Sturm (2007), Westenberger et al. (2006); 25. Ruvalcaba-Trejo and Sturm (2011), Westenberger et al. (2006); 26. Avila et al. (2003), Gerasimov et al. (2018), Maslov et al. (1994), Ruvalcaba-Trejo and Sturm (2011), Westenberger et al. (2006); 27. Blom et al. (1998); 28. Read, Fish, Muthiani, and Stuart (1993), Read, Jacob, Fish, Muthiani, and Stuart (1993); 29. Abraham, Feagin, and Stuart (1988), Benne et al. (1986), Bhat, Koslowsky, Feagin, Smiley, and Stuart (1990), Carnes et al. (2017), Corell, Myler, and Stuart (1994), Decker and Sollner-Webb (1990), Feagin, Abraham, and Stuart (1988), Feagin, Jasmer, and Stuart (1987), Feagin, Shaw, et al. (1988), Feagin and Stuart (1988), Kirby and Koslowsky (2017), Koslowsky, Bhat, Perrollaz, Feagin, and Stuart (1990), Read, Myler, and Stuart (1992), Read, Wilson, Myler, and Stuart (1994), Shaw et al. (1988), Simpson et al. (2017), Simpson et al. (2016), Souza et al. (1992), Souza, Shu, Read, Myler, and Stuart (1993); 30. Greif, Rodriguez, Reyna-Bello, Robello, and Alvarez-Valin (2015); 31. David et al. (2015); 32. Maslov and Simpson (1994), Lukeš et al. (1994); 33. Blom et al. (1998). The identities of the cultures described as *Crithidia* sp. KVI, *Herpetomonas* sp. TCC263 (Kolesnikov et al., 2003), and *W. inconstans* (Merzlyak et al., 2001) were not known and thus were not placed on the tree

TABLE 1 Comparison of analytical tools that have been created for transcriptomic analysis of products of U-indel genomes

Approach	PARERS	TREAT	*T-Aligner*
Citation	Carnes et al. (2017)	Simpson et al. (2016)	Gerasimov et al. (2018)
Open source (program)	No (unknown)	https://github.com/ubccr/treat (Go)	https://github.com/jalgard/T-Aligner3 (C++)
Input	Pooled library, consisting of individual libraries of PCR products of interest between ~180 and 300 nt, including Illumina indexing and sequencing adaptors	PCR products of target (gene) of interest, less than 550 bp including Illumina indexing and sequencing adaptors, with consistent 5′ and 3′ boundaries of editing	Library of cDNA fragments generated from poly(A)-enriched mitochondrial RNA, with insert sizes of 300–400 nt, although technically, PCR products could also be used as input
Templates required	Pre-edited and fully edited sequences for each amplified region of analysis	Pre-edited and fully edited sequences for each mRNA analyzed, and optionally, one or more alternatively fully edited sequences	DNA reference for each maxicircle-encoded mRNA or cryptogene, plus 50–100 nt flanking sequence on each reference
Sequencing strategy	50 nt single-end (MiSeq scale)	300 nt paired-end (MiSeq scale)	250 or 300 nt paired-end (MiSeq or HiSeq scale)
Output	Pooled output is separated by gene regions of interest, and then each gene's reads are placed into either pre-edited, canonically fully-edited, or partially edited categories. User then utilizes these read files with their own downstream analysis, with the partially edited category yielding the most telling information for mechanistic studies	Each read covers either the entire gene or a large portion of edited sequence. The pre-edited, junction, and canonically-edited regions of each read are identified. Population-wide measures of the junction regions (start and end sites, lengths, sequences) can be compared across samples by a web-based user interface producing constrainable histograms and tables of amplicon-specific data. Individual sequences can be viewed and compared to pre-edited and fully edited sequences. Data can be further analyzed using accompanying code written in R	Individually, reads for each cryptogene are mapped. The location of these mapped reads on the reference, and details about their editing are presented in multiple *T-Aligner* graphical outputs. Sequence file output includes reconstructed open reading frames (ORFs) in nucleotide and amino acid files, along with the level of read support for each reconstruction. Finally, editing pathways of up to three of the output ORFs can be mapped on the graphical output

Note. Lengths of PCR products or cDNA fragments used by TREAT or *T-Aligner* can be larger as sequencing technologies develop the capacity to generate longer reads.

about U-indel RNA editing. Newly available U-indel read population characterization methodologies have already enriched our understanding of these transcriptomes and will likely be used by more investigators in the future to answer distinct questions.

The dedicated tools developed to organize and describe U-indel mRNA transcriptomes are *T-Aligner* (Gerasimov et al., 2018), PARERS (Pipeline analyzing RNA editing RNA sequencing; Carnes et al., 2017), and TREAT (trypanosome RNA editing alignment tool; Simpson et al., 2016). The tools were designed to answer distinct questions, so each program has unique characteristics, outputs, and strengths (summarized in Table 1). Each of the three programs utilizes high throughput sequencing of cDNA, aligning or otherwise organizing sequences with highly variable U content. This is accomplished by masking the Ts of both reference and reads during mapping and retaining Ts during subsequent analysis. TREAT and *T-Aligner* are available as open source programs; PARERS is currently not a open source. Several other groups are using high throughput sequencing to collect and analyze small RNAs of the mitochondrial transcriptome; however, as they are not edited, their content has mainly been reported directly. No programs for analysis of small RNA populations have been published to date.

T-Aligner was designed to determine edited ORFs in previously unexplored kinetoplastids, and for comparison of mitochondrial transcriptomes across evolutionarily divergent Kinetoplastea members (Gerasimov et al., 2018). Of the three, *T-Aligner* is the only program that generates full-length final coding sequences de novo, solely from high throughput transcriptome reads. It is also the only program in which there is no polymerase chain reaction (PCR) amplification step, eliminating a potential source of bias. Reads in all states of editing are mapped to their reference cryptogenes obtained from maxicircle genomic DNA sequencing of the investigated organism. ORFs are then reconstructed from overlap graphs of read alignments using a breadth-first search algorithm. The longest ORFs starting at each possible start codon are collected. *T-Aligner* is the optimal tool in two instances. The first instance is when sequencing U-indel transcriptomes besides *L. tarentolae* and *T. brucei,* for which we lack the full complement of experimentally verified edited sequences. The second is in cases in which the experiment in question requires analysis of all of the mRNAs, rather than a few reporter transcripts.

Once the sequence of a cryptogene and at least one of its fully edited products is known, both TREAT and PARERS can be used to interrogate its editing parameters. PARERS was designed to interrogate the editing site specificities of the three subtypes of *T. brucei* RECCs (Carnes et al., 2017). PARERS permits examination of the first region of an mRNA that becomes edited (the most extreme 3′ edited region). The downstream analysis of sequences is not built into PARERS; methods to do so are user-generated. PARERS separates 50 nt single-end reads of PCR amplified pooled libraries into pre-edited, fully edited and "unmatched" groups, using a string match to regions bound by the gene-specific primers used to generate

the library, and segregate sequences by gene. The unmatched group contains all partially edited sequences, including those that do and do not contain junctions, which can be further examined by the user. The use of short reads and primers to many genes makes this program ideal for examining changes early in editing across many transcripts selected by the user.

TREAT was designed to assess the 3′ to 5′ progression of editing for a given transcript (Simpson et al., 2016). It determines reads that align to pre-edited, fully edited, or alternatively edited references. For partially edited sequences, it defines the extent of canonical fully edited sequence and the boundaries and characteristics of any junction present. Downstream analytical scripts allow the user to compare multiple parameters between samples such as (a) locations of intrinsic editing pause sites, (b) editing pause sites exacerbated in an experimental group compared to control, (c) the lengths, beginning and end sites, and sequences of junctions, and (d) plausible editing paths through sub-regions of a given transcript. These parameters can be evaluated relative to the positions of cognate gRNAs to generate mechanistic hypotheses (McAdams et al., 2018; Simpson et al., 2016, 2017). In contrast, *T-Aligner* is not able to assess where each read falls in the context of the full-length transcript sequence; thus, it reveals junction variation in aggregate with other variation but cannot examine specific junction sequence in the context of a full-length sequence. In summary, TREAT is the ideal tool to use when asking questions about underlying mechanisms of editing and when interrogating new factors for their roles in facilitating the initiation or progression of RNA editing.

4 | KNOWLEDGE GAINED FROM HIGH THROUGHPUT SEQUENCING METHODOLOGIES

4.1 | Features of the steady-state U-indel transcriptome

Early conventional sequencing studies hinted at several features of the kinetoplastid mitochondrial transcriptome that have been evaluated and largely confirmed in recent high throughput sequencing studies. The number and length of junction-containing sequences reported in studies utilizing conventional sequencing (Ammerman, Presnyak, Fisk, Foda, & Read, 2010; Benne et al., 1986; Bhat et al., 1990; Corell et al., 1994; Feagin, Abraham, & Stuart, 1988; Koslowsky et al., 1990, 1991, 1992; Read et al., 1992; Read, Wilson, et al., 1994; Souza et al., 1992, 1993) suggested that the presence of U-indel modifications other than those consistent with a canonical final product is extensive. Pairing these results with observations from full-gene PCR studies where all edited species were simultaneously amplified (Ammerman et al., 2011, 2013, 2010; Kumar et al., 2016; Schnaufer et al., 2001), it appeared that only a small fraction of most pan-edited mRNAs are fully edited. U-indel edited mRNAs derived from high throughput sequencing reads from *T. brucei*, *T. cruzi*, *Perkinsela*, and *Leptomonas pyrrhocoris* have since demonstrated this to be more universally true (Carnes et al., 2017; David et al., 2015; Gerasimov et al., 2018; McAdams et al., 2018; Simpson et al., 2016, 2017). Table 2 compares the percentage of sequences that were pre-edited, fully edited, partially edited containing junctions, and partially edited not containing junctions in some of these studies. However, given the variation in techniques used and types of sequences reported (such as partially edited sequences only), the numbers presented are not directly comparable. Further studies must be initiated to establish the universality of features of U-indel editing shown in Table 2.

High throughput analyses of the mitochondrial transcriptomes of *L. pyrrhocoris* and *T. cruzi* revealed insights to the degree of editing of a transcriptome, in other words, the fraction of the total mRNA population for each specific transcript that is edited in any way, and how this varies across divergent members of the order Kinetoplastida (Gerasimov et al., 2018). Sequence analysis revealed a high variability in the degree of editing between (a) transcripts, (b) the 5′ and 3′ regions of a single transcript, and (c) organisms. In *L. pyrrhocoris*, the *COIII* mRNA population is almost completely fully edited (>95%) at its tiny three-site-edited domain (Gerasimov et al., 2018). Pan-edited *L. pyrrhocoris* mRNAs mainly exhibit an expected profile, wherein a majority of reads are edited at sites contained in the extreme 3′ end of the edited region, with a typical drop-off of editing towards the 5′ end of the transcript. Because the 5′ end of an mRNA is the last part to be edited, the percentage of reads edited at the 5′-most sites can be taken as an approximate measure of the fully edited mRNA fraction. For *L. pyrrhocoris* pan-edited mRNAs, fully edited molecules range from only a few reads (*ND*4 and *G4*) to ~50% of the population (*RPS12*) (Gerasimov et al., 2018). The transcript-by-transcript variability in degree of editing may be matched by variability across organisms. When *T-Aligner* output from *T. cruzi RPS12* and *COIII* are compared with that of *L. pyrrhocoris*, it is clear that a much higher fraction of reads from *L. pyrrhocoris* 5′ regions contain edited sequence. In fact, for dividing epimastigote stage *T. cruzi*, barely any full-length *RPS12* and *COIII* mRNA molecules are translatable (Gerasimov et al., 2018). This is also the case for procyclic stage *T. brucei*, as Simpson et al. (2016) found less than 6% of *RPS12* and *ND7* 5′ edited domain cDNA reads to be canonically fully edited within their ORFs. However, in a different life stage or environment, *T. cruzi* or *T. brucei* editing of certain mRNA may be more robust (above; Shaw, Kalem, & Zimmer, 2016). In summary, the degree to which a transcript population achieves fully edited status is highly variable by transcript, life stage, and organism, and can now be quantified. What such differences mean at the protein or functional level is a challenging question yet to be elucidated.

TABLE 2 Comparison of relative number of sequences obtained from cryptogenes at various stages of editing completion

Reference	Organism	Method	Gene	Total sequences	Fully edited (%)	Partially edited (%)	Partially edited, no junction (%)	Partially edited, with junctions (%)	Pre-edited (%)
Koslowsky et al. (1991)	*Trypanosoma brucei*	Conventional cloning and sequencing	*ND7*	30[a]	0 (0)	30 (100)	4 (13)	26 (87)	0 (0)
Koslowsky et al. (1991)	*T. brucei*	Conventional cloning and sequencing	*A6*	23[a]	0 (0)	23 (100)	2 (9)	21 (91)	0 (0)
Maslov et al. (1992)	*Leishmania tarentolae*	Conventional cloning and sequencing	*RPS12*	26[a]	0 (0)	25 (96)	12 (46)	14 (54)	1 (4)
Landweber et al. (1993)	*Herpetomonas megaseliae*	Conventional cloning and sequencing	*COIII*	36[a]	0 (0)	36 (100)	15 (42)	21 (58)	0 (0)
Landweber et al. (1993)	*Lafontella mariadeanei*	Conventional cloning and sequencing	*COIII*	25[a]	0 (0)	25 (100)	2 (8)	23 (92)	0 (0)
Maslov and Simpson (1994)	*Trypanoplasma borreli*	Conventional cloning and sequencing	*CYb*	19[a]	0 (0)	19 (100)	2 (10)	17 (90)	0 (0)
Maslov et al. (1998)	*Phytomonas serpens*	Conventional cloning and sequencing	*RPS12*	4	4 (100)	0 (0)	0 (0)	0 (0)	0 (0)
Maslov et al. (1999)	*P. serpens*	Conventional cloning and sequencing	*A6*	7	7 (100)	0 (0)	0 (0)	0 (0)	0 (0)
Maslov et al. (1999)	*P. serpens*	Conventional cloning and sequencing	*ND8*	16	0 (0)	10 (66)	n/a	n/a	6 (34)
Maslov et al. (1999)	*P. serpens*	Conventional cloning and sequencing	*CR3/G3*	4	3 (75)	1 (25)	n/a	n/a	0 (0)
Ammerman et al. (2010)	*T. brucei*	Conventional cloning and sequencing	*RPS12*	51	6 (12)	40 (78)	6 (15)	34 (85)	5 (10)
David et al. (2015)	*Perkinsela*	High throughput + *T-Aligner* precursor	*COI 5′ region*	823	548 (67)	157[b] (19)	n/a	n/a	118 (14)
David et al. (2015)	*Perkinsela*	High throughput + *T-Aligner* precursor	*A6*	3,755	883 (23)	1979[b] (53)	n/a	n/a	893 (24)
David et al. (2015)	*Perkinsela*	High throughput + *T-Aligner* precursor	*RPS12*	89	45 (51)	44[b] (49)	n/a	n/a	0 (0)
Simpson et al. (2016)	*T. brucei*	High throughput + TREAT	*RPS12*	251,006	14,385 (6)	201,170 (80)	9179 (4)	191,991 (77)	35,451 (14)
Simpson et al. (2016)	*T. brucei*	High throughput + TREAT	*ND7 5′*	798,405	46,488 (6)	502,591 (63)	25,258 (3)	477,333 (60)	249,326 (31)
Carnes et al. (2017)	*T. brucei*	High throughput + PARERS	*ND7 5′*	252,105[a,c] (WT)	0 (0)	109,613 (43)	n/a	≥50,031 (20)	142,492 (56)
Carnes et al. (2017)	*T. brucei*	High throughput + PARERS	*A6*	66,513[a,c] (WT)	0 (0)	1955 (03)	0 (0)	1955 (100)	64,558 (97)

Note. Sequences obtained that reflect a translatable, edited consensus from start codon to stop codon are listed here as "fully edited," regardless of whether there are editing differences in the 5' untranslated region. Except for sequences obtained in Simpson et al. (2016), and for *Phytomonas serpens*, sequences presented in this table largely reflect fragments of an entire mRNA. Therefore, the data here are not meant to reflect exact percentages of molecules in various stages of the editing process. Rather, they reflect variation between organisms, and in the case of *T. brucei* sequencing, consistency regardless of technical approach. Only sequencing endeavors for which relative numbers of sequences in various editing states were reported were considered for this table.

[a] In these studies, only partially edited sequences were examined.

[b] In the *Perkinsela* study, reads were described as containing at least one instance of alternative editing; we placed these in the junction-containing category.

[c] Only the first few editing sites were analyzed.

4.2 | Insights into the generation and roles of junction regions and noncanonically edited mRNA

It is convenient to call editing a "noisy" process for lack of a more concise description for the surprisingly high ratio of noncanonical relative to canonical U-indels observed in the majority of pan-edited molecules. However, this terminology masks the evidence that in some or most cases, editing of a site or sites in a manner that does not match the canonical sequence may play a functional role.

Summarized in Figure 3 are three ways that noncanonical modifications may serve a functional role:

1. Their presence may mark dead-end products of the editing process. However, even in this case, they can be used as a point of regulation if the organism can adjust their abundance relative to canonically edited products as needed.
2. They are essential intermediates in the progression of editing.
3. Their presence will result in alternative product(s) in addition to the canonical product.

Results of high throughput studies are clarifying the merits and limits of these hypotheses.

Although noncanonical U insertions and deletions have been found within, rather than at, the leading edge of edited sequences in *Perkinsela*, *L. pyrrhocoris*, and *T. brucei* (David et al., 2015; supplemental material of Gerasimov et al., 2018; Simpson et al., 2016), they are predominantly found at the leading 5′ edge of editing, in junctions. This is evidenced by previously described individual cloning and sequencing studies (including those of Table 2), recent high throughput sequencing studies in *T. brucei*, and in the multitude of 5′-truncated reconstructed ORFs with misediting within a short initial sequence length in *L. pyrrhocoris*, *Perkinsela,* and *T. cruzi* (David et al., 2015; Gerasimov et al., 2018; Simpson et al., 2016). That mRNAs containing junctions comprise the majority of partially edited mRNAs supports some functional role for these noncanonical editing modifications (Table 2; Ammerman et al., 2010; Koslowsky et al., 1991).

It is especially challenging to discern whether a junction containing partially edited sequence is either a dead-end product or an intermediate (first and third options in Figure 3). Even with the ability to obtain many sequences, it is currently not possible to kinetically track the progression of editing of a single transcript in real time. Thus, the fate of any partially edited molecule, whether to undergo future editing remodeling events or else be degraded, is unknown. However, high throughput analysis of sequences containing given misedited segments relative to other sequences in the population may reveal those that are more suggestive of one or the other scenario (Figure 4). For example, a likely profile of a "dead-end" product is a partially edited transcript with a junction containing a series of modifications that could not be productively edited without extensive reversal of multiple existing modifications (Figure 4, yellow arrows). Conversely, an intermediate product would be identifiable by the presence of other junctions that could be generated from it by single modifications that resolve to a conserved edited sequence (Figure 4, yellow box).

There is a lack of consensus regarding the nature of junctions as intermediates. Authors of the *Perkinsela* U-indel transcriptome sequencing projects noted that the vast majority of junction-containing reads (containing noncanonical U insertions or deletions) were present in extremely low copy number. Assuming that junction sequences that are part of a regular pathway of successive editing modifications and mRNA/gRNA realignments would appear with a certain frequency, they argued that their findings indicated that the mRNAs containing such junctions were likely dead-end products (David et al., 2015). Furthermore, high abundances of reads present at single or low copy number that contained alternative sequence were also found in the *L. pyrrhocoris* and *T. cruzi* transcriptomes (Gerasimov et al., 2018). The authors here also noted the previously described situation of too many junctions requiring multiple revisions to realistically envision them as intermediates. It is possible that the fraction of junctions representing editing intermediates may differ by species. To resolve this issue, more quantitative analyses of these non-PCR-based data specifically to address this question is needed, such as quantifying variation within junction

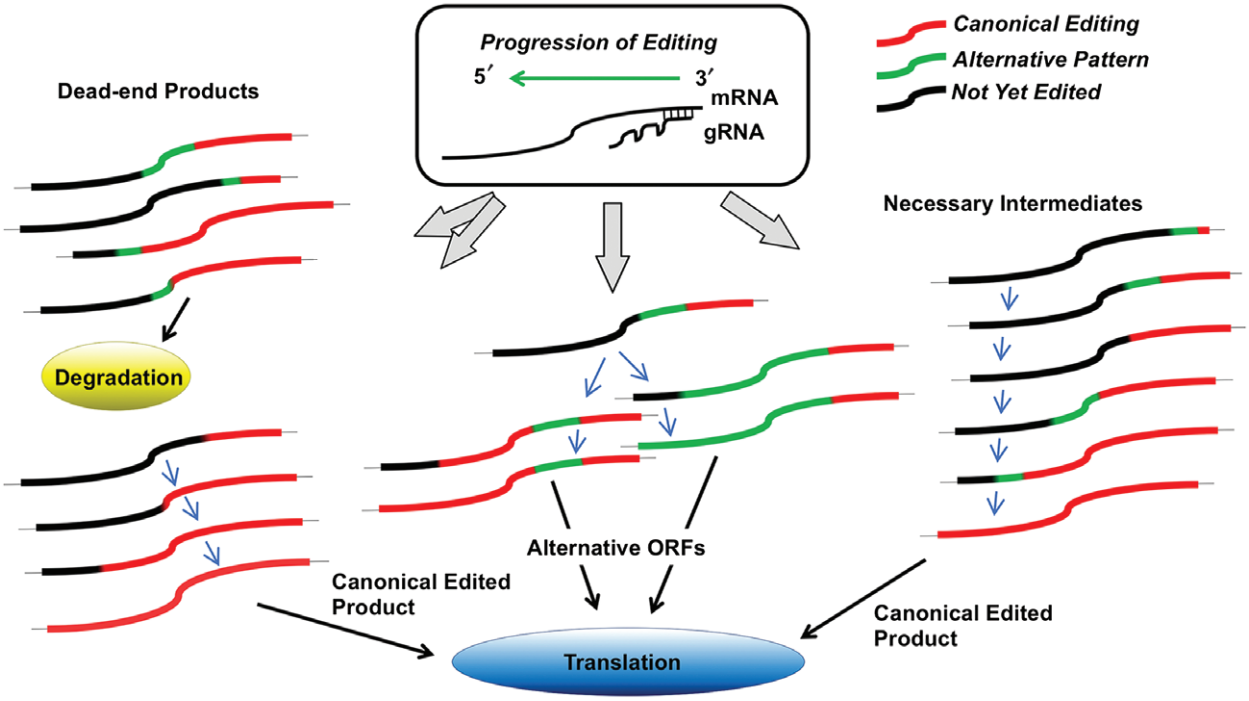

FIGURE 3 Reasons for and potential fates of mRNA molecules that exhibit alternative (non-canonical) editing patterns. Arrows represent progression of editing in a single molecule that has been color-coded to indicate its editing state. In the first panel, the alternative sequences are present only in dead-end edited products destined for degradation. In the second panel, the alternative sequences are part of a product with a region that is alternatively edited to generate an alternative open reading frame (ORF). In the third panel, the region with alternative editing is present in junctions that will continue to be edited until the canonical sequence is obtained

sequences and determining the frequency at which ORFs contain internal misedited sequence. Currently, such analyses are outside the scope of the *T-Aligner* software.

The hypothesis of junctions as intermediates has been investigated primarily in *T. brucei.* Authors of two other studies also observed great diversity in junction sequences; however, these modifications were not random as would be expected with mere "noise" but rather occurred at a subset of editing sites suggesting a directed process (Carnes et al., 2017; Simpson et al., 2016). Analysis of these sequences showed many junctions that differed by single editing modifications as expected with successive intermediates (Ammerman et al., 2010; Koslowsky et al., 1992; McAdams et al., 2018; Simpson et al., 2016, 2017). Furthermore, depletion of certain RESC components that impede 3′ to 5′ editing progression simultaneously resulted in a loss of junction sequences (Simpson et al., 2017), suggesting that junction sequences are necessary for normal editing. The authors point to the ever-shifting mRNA/gRNA duplex as a potential source for thermodynamically unavoidable intermediates that could direct noncanonical modifications at consistent locations, generating junctions. The location of junctions at the 5′ leading edge of editing is consistent with this model. Expansion of order of modification analysis (Figure 4), especially under stress conditions, different life cycle stages and during depletion of editing factors will be needed to confirm these initial hypotheses.

Evidence that noncanonical editing generates alternative ORFs, as shown in the middle panel in Figure 3, has been extensively sought with some success. Alternatively edited mRNA molecules or reconstructed ORFs have been identified in *T. brucei*, *T. cruzi*, *L. mexicana*, *L. pyrrhocoris*, and *Perkinsela* (Gerasimov et al., 2018; Kirby & Koslowsky, 2017; Madina et al., 2014; Maslov, 2010; Ochsenreiter & Hajduk, 2006; Read et al., 1992; Read, Wilson, et al., 1994; Simpson et al., 2016). The more targeted studies have also identified the gRNAs that could direct these alternative editing events (Kirby & Koslowsky, 2017; Madina et al., 2014; Ochsenreiter & Hajduk, 2006). In the most complete pursuit of such products, identification of an alternatively edited mRNA, a potential causative gRNA, and antibody-detectable protein product arising from the *T. brucei COIII* cryptogene were identified (Ochsenreiter, Anderson, Wood, & Hajduk, 2008; Ochsenreiter & Hajduk, 2006). The abundance of this product, identified by conventional sequencing, is not known. Ideally, in addition to the types of evidence collected for *T. brucei COIII*, proof of alternative editing would also include proof of ribosome association of the alternative transcript, and that the encoded protein (identified by mass spectrometry or immunodetection) is depleted upon disruption of U-indel RNA editing. This idealized scenario faces technical hurdles. Approaches for ribosome profiling of organellar ribosomes have yet to be developed, and even protein products predicted from canonically edited mRNAs have been notoriously difficult to observe by traditional methods, possibly due to their extreme hydrophobicity (Acestor et al., 2011; Horváth, Berry, & Maslov, 2000; Horváth, Kingan, & Maslov, 2000; Panigrahi et al., 2009; Škodová-Sveráková, Horváth, & Maslov, 2015). As even the presence of a protein may not be evidence that it is functional (Ziková, Verner, Nenarokova,

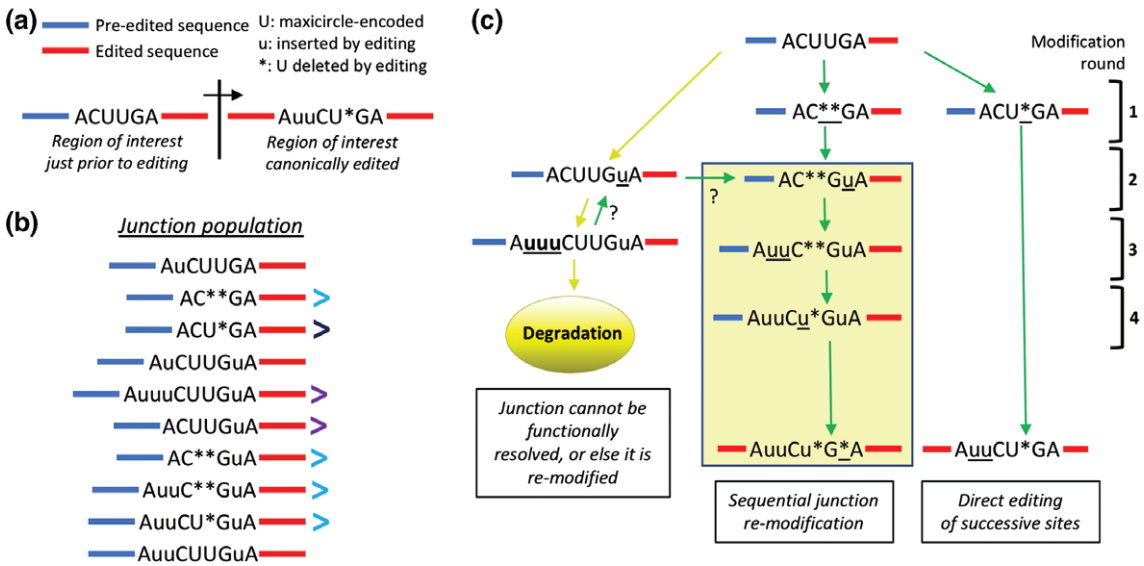

FIGURE 4 Successive editing modifications successively convert junction sequence to canonically edited sequence. (a) The pre- and fully edited versions of a short model sequence, found within an mRNA requiring editing, are shown. (b) Example junction sequences found in this region such as those observed through high throughput sequencing. The dark blue and light blue, and violet > symbols represents editing intermediates shown in c. (c) Noted junction sequences from b have been ordered such that a progression from one junction to another can be generated by a single editing modification (underlined) from the sequence before it, beginning with the pre-edited and ending with the fully edited sequence. On the far left is a pathway where successive noncanonical modifications progressively build up, resulting in a junction that possibly can no longer be productively re-modified (delineated with bold u insertions). On the right, two paths of editing are shown, one with junctions resolving to the fully edited sequence (left) and the other showing a direct progression of editing (right). These pathways are based on observations in Simpson et al. (2017)

Michels, & Lukeš, 2017), investigations of potential product function, such as those performed by Ochsenreiter, Anderson, et al., 2008, would ultimately be necessary. Development of tools for genetic manipulation of the trypanosome mitochondrion would greatly aid these endeavors.

However, with the advent of high throughput sequencing, examples of reading frames generated by alternative editing are compounding. Small alternative editing events that result in frameshifts, particularly near the 5' end of the editing domain, would cause the formation of an alternative N-terminus or else an ORF with an entirely different sequence. These have been identified or inferred in *T. brucei* (Kirby & Koslowsky, 2017; Koslowsky et al., 2014; Madej, Niemann, Hüttenhofer, & Göringer, 2008; Madina et al., 2014; Ochsenreiter, Cipriano, & Hajduk, 2008; Simpson et al., 2016). The noncanonical *L. pyrrhocoris* and *T. cruzi* full length ORF reconstructions with highest support were those in which one or a few alternative editing events in a discrete region shift the reading frame such that the encoded protein contains a canonical central domain with an altered N- or C-terminus (Gerasimov et al., 2018). In fact, this is what was observed to generate the *T. brucei* COIII alternative (Ochsenreiter & Hajduk, 2006). Additionally, as represented by the ORF primarily colored green in the middle panel of Figure 3, *T-Aligner* reconstructed rare full-length ORFs (that utilized only one or very few reads over the divergently edited region), with long stretches of alternatively edited sequence that encodes an entirely different amino acid sequence. These regions were extensive enough that multiple gRNAs, up to seven, would potentially be required to direct their editing. The ORFs were hypothesized to be too rare to be functionally relevant, but revealed an inherent capacity of the genome to produce entirely different proteins (Gerasimov et al., 2018). The mounting evidence of alternative RNA products will undoubtedly spur the necessary technical advances to determine whether the fate of alternatively edited mRNAs is in fact translation. Additionally, with methods now in place to acquire complete expressed gRNAs populations (Kirby et al., 2016; Koslowsky et al., 2014), abundances of gRNAs responsible for alternatively edited transcripts can be compared to their canonically guiding counterparts. These ratios in turn can be compared to the distribution of canonical and alternatively edited sequences in U-indel transcriptomic data.

4.3 | High throughput sequencing as a tool to illuminate mechanism

The impact of high throughput sequencing to enhance mechanistic studies of editing, going beyond protein–protein interactions, is already evident. Studies utilizing this technology both illuminate the nature of kinetoplastid mitochondrial transcriptomes and reveal aspects of their generation. For example, the catalytic RECCs (Figure 1b) are present in three predominant types, each containing a distinct endonuclease (KREN1, KREN2, or KREN3), which cleaves mRNA prior to U insertion or deletion. Using high throughput sequencing and the PARERS tool, Carnes et al. (2017) investigated the in vivo specificity of KRENs by analyzing editing modifications across multiple RNAs in *T. brucei* cells with one, two or all three KRENs deleted. These studies showed that KREN2 and KREN3 cleave at U insertion sites, with differing but overlapping specificities, and that KREN1 exhibits specificity for U deletion sites. In addition, while distinct RECCs exhibited specific preferences in the types of sites at which their component endonucleases cleaved RNA, the number of Us inserted and deleted at a given location varied widely. These findings suggest that noncanonical U indels within junctions are unlikely to be a product of aberrant RECC action, but rather reflect variable modifications at locations that are targeted with some specificity.

Simpson et al. (2017) investigated the functions of three components of the REMC module of RESC (Figure 1b). While TbRGG2, MRB8180, or MRB8170/4160 were known to interact in vivo (Ammerman et al., 2012; Aphasizheva et al., 2014; Hashimi, Cicova, Novotna, Wen, & Lukeš, 2009; Kafková et al., 2012; Panigrahi et al., 2008; Weng et al., 2008), their precise relationships and functions were not understood. TREAT analysis revealed that the three proteins have distinct but overlapping functions in facilitating progression of editing along a transcript. TREAT also allowed confirmation, by a different methodology, of the discovery (Dixit et al., 2017) that MRB8170/4160 acts in editing initiation. Furthermore, RPS12 mRNA partially edited sequences were analyzed in a stepwise manner to determine likely paths of editing modification in the region of editing directed by the first gRNA. In wild type cells, sequences containing or lacking junctions in this region were present almost equally. Depletion of either TbRGG2 or MRB8180 dramatically decreased the proportion of junction-containing mRNAs. That is, integral components of the editing machinery actually facilitate the production of noncanonically edited junctions, while at the same time causing a fatal editing defect. This study clarified the individual roles of members of the REMC subcomplex and provides the strongest evidence to date that junctions are essential for the progression of editing.

McAdams et al. (2018) used TREAT analysis to explore a previously unstudied component of RESC, MRB7260, and showed editing pausing patterns in MRB7260 knockdowns that were very distinct from those of other RESC components (Simpson et al., 2017). The impacts of silencing, including grouped pauses just beyond gRNA anchor sequences, several pauses at gRNA ends, and apparent utilization of an incorrect gRNA, could also be determined by these methods. Parallel co-immunoprecipitation, glycerol gradient, and RNA immunoprecipitation studies supported a role for MRB7260 in gRNA trafficking and utilization. These studies demonstrate how high throughput analysis of editing defects in specific knockdowns support and extend traditional genetic approaches.

Two studies have employed high throughput sequencing for the characterization of RNA populations associated with the editing machinery. In the very first utilization of high throughput sequencing in a study of U-indel RNA editing, Madina et al. (2011) demonstrated differences in the gRNA populations isolated from native complexes containing RNA Editing Helicase 2 compared with those of the RESC component MRB3010. However, the small amount of RNA present in immunoprecipitated complexes yielded incomplete gRNA coverage and low read counts for some gRNAs. Subsequently, Dixit et al. (2017) combined in vivo mRNA-protein crosslinking with high throughput sequencing to define the cohort of *T. brucei* mitochondrial RNAs bound by MRB8170 and MRB4160. These authors demonstrated the feasibility of obtaining 50,000–100,000 reads from a single affinity purification of a tagged mitochondrial protein. Their experiments highlighted differences in mRNA preferences of these two orthologous RNA binding proteins, activities which had otherwise been difficult to distinguish. However, the authors only examined reads that matched the pre-edited or fully edited sequences, leaving the large set of alternatively edited or partially edited reads uninvestigated. We expect that with the advent of approaches that utilize T-masking to align sequences and the growing library of partially edited intermediates as a reference, this "missing" data can be utilized and may yield additional insights.

Finally, sequencing of minicircle transcriptomes can inform the mechanisms of gRNA processing and contribute to models of U-indel editing. Suematsu et al. (2016) analyzed high throughput sequences to identify minicircle-derived RNAs transcribed in antisense orientation and develop a model for the processing and maturation of gRNAs. Recently published high throughput transcriptomes of gRNAs from both procyclic and bloodstream form *T. brucei* generated nearly complete gRNA coverage for most mRNAs (Kirby et al., 2016; Koslowsky et al., 2014), thereby permitting analysis of this U-indel mRNA transcriptome in the context of its corresponding gRNA population. Structural interactions between gRNAs and mRNAs can be predicted, and alternative and/or rare gRNAs can be observed that may be responsible for certain junction region sequences or alternatively edited products. For instance, some junction-containing mRNAs likely result from editing guided by a noncognate gRNA that has spuriously anchored at an aberrant location (as was shown for *L. tarentolae COIII* and *RPS12* in Sturm, Maslov, Blum, and Simpson (1992)). High throughput methods can now tell us how common this junction-generating mechanism likely is. Already, high throughput maxicircle transcriptome output in *T. brucei* has been directly compared to gRNA coverage maps to support models of proteins that affect either progression through a gRNA or gRNA exchange (Kirby & Koslowsky, 2017; McAdams et al., 2018; Simpson et al., 2016, 2017). Moreover, early studies suggesting that gRNAs are consumed during editing were confirmed by Aphasizheva et al. (2014) by using high throughput sequencing to interrogate the gRNA population in cells in which editing was inhibited. Clearly, the relationship between the small RNA and maxicircle U-indel edited transcriptomes has great potential to reveal dynamics and variation in gRNA/mRNA interactions.

5 | HIGH THROUGHPUT SEQUENCING TO ILLUMINATE U-INDEL EDITING'S ORIGINS AND ADVANTAGES

The complexity of the U-indel editing process combined with its apparent confinement to one organelle in one group of organisms has fueled multiple ideas about why it exists in nature. There are multiple and not necessarily competing hypotheses regarding the origins of editing. Already, authors of studies utilizing high throughput sequencing have begun to weigh in on these possibilities (Gerasimov et al., 2018; Kirby & Koslowsky, 2017). One model postulates that the entire system emerged despite the lack of any adaptive advantage of the U-indel editing process. Under the constructive neutral evolution scenario, existing and then increasingly specialized machinery emerged to counter an original mutation followed by subsequent ones. Since the likelihood of mutation back to an original nucleotide is very low relative to that of a new mutation, the editing machinery became essential (Gray, 2012; Gray, Lukeš, Archibald, Keeling, & Doolittle, 2010; Lukeš, Archibald, Keeling, Doolittle, & Gray, 2011; Stoltzfus, 1999).

Selective advantages of possessing U-indel editing have also been postulated. These selective advantages could root its origins, or at least some of its development to its present state among kinetoplastids, in natural selection. The ability to generate alternative products from a single protein-coding locus discussed above is one of these. Another is that U-indel editing could serve as a mechanism of change to more rapidly evolve mitochondrial genome output (Landweber & Gilbert, 1993), leading to greater "evolvability" for an organism. Its descendants will be more successful although the mechanism in question has no immediate benefit (Gommans, Mullen, & Maas, 2009; Speijer, 2008). Establishing an additional regulatory level with which to manipulate gene expression, or a level at which to fine-tune expression control, is also of potential benefit, especially across a varied life cycle (Landweber, 1992; Speijer, 2008). Another perhaps less intuitive benefit of editing is that the fragmentation of genetic information resulting from gRNAs encoded on separate molecules from the mRNA cryptogene may have a protective effect for the genome as a whole under periods of relaxed selection. For example, this could be beneficial in the *T. brucei* bloodstream stage, where the organism requires few mitochondrial-encoded proteins (Speijer, 2008). In fact, fragmentation has been proposed to be mathematically advantageous for maintaining genome integrity during periods of relaxed

selection (Buhrman, van der Gulik, Severini, & Speijer, 2013). Finally, a situation may have arisen in which U-indel editing was necessary to counteract another evolutionary development. It has been proposed that G-quadruplexes inherent in the maxicircle genome are important to separate its replication from its transcription. In this case, G nucleotide regions of the pre-edited mRNA quadruplexed with the G nucleotides in the DNA acting as the switch between transcription and replication (Leeder, Hummel, & Göringer, 2016).

The origins and evolution of editing are entwined with the possibility that organisms are currently exploiting editing as a regulatory mechanism, a scenario in which it is possible to point to actual examples. As described above, that *T. brucei* edits certain mRNAs in one or another of its life stages has long been known (and recently systematically shown in Gazestani, Hampton, Shaw, Salavati, and Zimmer, (2018)). Mimicking *T. cruzi* starvation within the insect gut in culture results in an increase in edited but not pre-edited forms of analyzed mitochondrial mRNAs (Shaw, Kalem, & Zimmer, 2016), an example of a regulatory response to an environmental stimulus. A broader understanding of such advantages of editing, as well as the diversity of editing parameters among kinetoplastids, such as the degree of editing within cryptogene loci (Aravin et al., 1998; Blom et al., 1998; Gao, Kapushoc, Simpson, Thiemann, & Simpson, 2001; Gerasimov et al., 2012; Landweber et al., 1993; Landweber & Gilbert, 1993; Maslov, 2010; Maslov et al., 1994; Merzlyak et al., 2001; Ramírez et al., 2011; Thiemann et al., 1994), will inform our perspective of U-indel editing's origins and evolution.

Currently, probing models of selective advantage and/or constructive neutral evolution of U-indel editing is limited by the disproportionate use of only a few selective kinetoplastids as an investigative focus. However, the genome revolution is allowing the discovery and characterization of new kinetoplastids (David et al., 2015; Flegontov et al., 2013, 2016), of which there are hundreds or more (Maslov, Votýpka, Yurchenko, & Lukeš, 2013). Once a kinetoplastid's maxicircle genome is sequenced, it will be increasingly simple to obtain its mitochondrial U-indel and gRNA transcriptomes, even if investigation of the editing complexes and their functions remain labor-intensive.

Specific examples of questions to be addressed are easy to identify. For instance, loss of some editing and minicircle complexity in *L. tarentolae* cultured for long time periods relative to a recent isolate was reported (Gao et al., 2001; Thiemann et al., 1994), but extensively cultured *L. mexicana* did not exhibit similar loss (Maslov, 2010). With our new sequencing capabilities, this phenomenon can be further explored to determine whether loss of editing is typical for trypanosomatid species in culture, or rather reflects an extensive natural variation in editing and gRNA complexity (or transcription) among species and isolates. Similarly, comparisons of editing between the closely related *T. brucei, T. congolense*, and *T. vivax* could be performed at a more informative level (Greif et al., 2015; Kirby & Koslowsky, 2017; Read, Fish, et al., 1993). Use of high throughput sequencing when comparing editing between environmental circumstances and life stages will eliminate the bias of examining only easy-to-analyze genes, and it may generate data of a depth that will illuminate the underlying mechanisms for these differences. Examination of more kinetoplastid genomes will reveal how extensively alternative editing resulting in multiple protein products is utilized, and whether this varies by species.

Finally, it may be possible to ultimately tease apart the impetus for U-indel editing's emergence as being more rooted in selection or in constructive neutral evolution with the aid of these new technologies. Criteria have been proposed to test processes that result from constructive neutral evolution as opposed to selection. For instance, U-indel editing resulting from constructive neutral evolution might be expected to result in a smooth increase in complexity over time in descendants, and its distribution should not be too varied (Speijer, 2011). These sorts of comparisons can be facilitated with high throughput sequencing of multiple maxicircle and minicircle genomes and transcriptomes. In summary, the sequencing revolution has positioned us to answer not only important mechanistic but also fundamental questions about the role of U-indel editing across the wide range of species that utilize this fascinating RNA modification process.

6 | CONCLUSION

The application of high throughput sequencing technologies to U-indel editing holds the promise of rapid and fundamental advances in our understanding of numerous aspects of this complex process, and recent publications from several laboratories demonstrate the power of this approach. Questions that can now be addressed and features whose functional significance can now be understood using these technologies and platforms include:

1. What percentage of a given transcript population is fully or partially edited, and how does this parameter differ between transcripts?
2. How does editing progress along different mRNA populations? What are the intrinsic barriers to editing progression? How does intra-mRNA or mRNA–gRNA structure contribute to progression?

3. How prevalent are misedited junctions, and what are their features? What fraction of junctions exhibits sequences that could be easily re-modified to canonical edited sequence and what fraction spans long distances (perhaps greater than one gRNA length) that appear to be difficult to properly re-modify?
4. How prevalent are sequences that apparently encode alternative ORFs?
5. What is the extent of U-indel editing-generated 5′ untranslated region diversity? Future studies could also address whether this parameter correlates with ribosome association.

Importantly, analysis of the above parameters in different life cycle stages or environmental conditions will reveal the degree to which the editing process is regulated and the steps that are amenable to regulation. High throughput sequencing-based examination of cells depleted of specific editing machinery subunits or expressing mutant subunits will provide key insights into the mechanism of U-indel editing. Finally, comparison of the above features between mRNA populations from different kinetoplastids will illuminate the evolution, and perhaps the *raison d'etre* of a bizarre and fascinating process.

ACKNOWLEDGMENTS

The authors would like to thank Vyacheslav Yurchenko, University of Ostrava, Ostrava, CZE, for assistance in discerning the identities of analyzed kinetoplastids presented in Figure 2, and Julius Lukeš, Czech Academy of Sciences, for critical reading of the manuscript. Final editing by members of the Zimmer lab was also appreciated. This work was supported by the American Heart Association (16SDG26420019 to S.L.Z.) and National Institutes of Health (AI061580 to L.K.R.).

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

ORCID

Sara L. Zimmer http://orcid.org/0000-0002-1707-1839

Rachel M. Simpson http://orcid.org/0000-0003-2330-3410

REFERENCES

Abraham, J. M., Feagin, J. E., & Stuart, K. (1988). Characterization of cytochrome c oxidase III transcripts that are edited only in the 3′ region. *Cell*, *55*(2), 267–272. https://doi.org/10.1016/0092-8674(88)90049-9

Acestor, N., Zikova, A., Dalley, R. A., Anupama, A., Panigrahi, A. K., & Stuart, K. D. (2011). *Trypanosoma brucei* mitochondrial respiratome: Composition and organization in procyclic form. *Molecular & Cellular Proteomics*, *10*(9), M110.006908. https://doi.org/10.1074/mcp.M110.006908

Ammerman, M. L., Downey, K. M., Hashimi, H., Fisk, J. C., Tomasello, D. L., Faktorova, D., … Read, L. K. (2012). Architecture of the trypanosome RNA editing accessory complex, MRB1. *Nucleic Acids Research*, *40*(12), 5637–5650. https://doi.org/10.1093/nar/gks211

Ammerman, M. L., Hashimi, H., Novotna, L., Cicova, Z., McEvoy, S. M., Lukeš, J., & Read, L. K. (2011). MRB3010 is a core component of the MRB1 complex that facilitates an early step of the kinetoplastid RNA editing process. *RNA*, *17*(5), 865–877. https://doi.org/10.1261/rna.2446311

Ammerman, M. L., Presnyak, V., Fisk, J. C., Foda, B. M., & Read, L. K. (2010). TbRGG2 facilitates kinetoplastid RNA editing initiation and progression past intrinsic pause sites. *RNA*, *16*(11), 2239–2251. https://doi.org/10.1261/rna.2285510

Ammerman, M. L., Tomasello, D. L., Faktorova, D., Kafková, L., Hashimi, H., Lukeš, J., & Read, L. K. (2013). A core MRB1 complex component is indispensable for RNA editing in insect and human infective stages of *Trypanosoma brucei*. *PLoS One*, *8*(10), e78015. https://doi.org/10.1371/journal.pone.0078015

Aphasizheva, I., & Aphasizhev, R. (2016). U-insertion/deletion mRNA-editing holoenzyme: Definition in sight. *Trends in Parasitology*, *32*(2), 144–156. https://doi.org/10.1016/j.pt.2015.10.004

Aphasizheva, I., Zhang, L., Wang, X., Kaake, R. M., Huang, L., Monti, S., & Aphasizhev, R. (2014). RNA binding and core complexes constitute the U-insertion/deletion editosome. *Molecular and Cellular Biology*, *34*(23), 4329–4342. https://doi.org/10.1128/MCB.01075-14

Aravin, A. A., Yurchenko, V., Merzlyak, E. M., & Kolesnikov, A. A. (1998). The mitochondrial ND8 gene from *Crithidia oncopelti* is not pan-edited. *FEBS Letters*, *431*(3), 457–460. https://doi.org/10.1016/s0014-5793(98)00813-8

Avila, A. R., Dallagiovanna, B., Yamada-Ogatta, S. F., Monteiro-Goes, V., Fragoso, S. P., Krieger, M. A., & Goldenberg, S. (2003). Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis. *Genetics and Molecular Research*, *2*(1), 159–168.

Benne, R., Van den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H., & Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*, *46*(6), 819–826. https://doi.org/10.1016/0092-8674(86)90063-2

Bessolitsyna, E. A., Fediakov, A. V., Merzliak, E. M., & Kolesnikov, A. A. (2005). ATPase subunit 6 gene of *Leptomonas seymouri* (Trypanosomatidae) is transcribed and edited as a polycistronic transcript. *Molekuliarnaia Biologiia*, *39*(1), 61–66. https://doi.org/10.1007/s11008-005-0007-y

Bhat, G. J., Koslowsky, D. J., Feagin, J. E., Smiley, B. L., & Stuart, K. (1990). An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. *Cell*, *61*(5), 885–894. https://doi.org/10.1016/0092-8674(90)90199-o

Bhat, G. J., Myler, P. J., & Stuart, K. (1991). The two ATPase 6 mRNAs of *Leishmania tarentolae* differ at their 3′ ends. *Molecular and Biochemical Parasitology*, *48*(2), 139–149. https://doi.org/10.1016/0166-6851(91)90110-r

Blom, D., de Haan, A., van den Berg, M., Sloof, P., Jirku, M., Lukeš, J., & Benne, R. (1998). RNA editing in the free-living bodonid *Bodo saltans*. *Nucleic Acids Research*, *26*(5), 1205–1213. https://doi.org/10.1093/nar/26.23.iii

Buhrman, H., van der Gulik, P., Severini, S., & Speijer, D. (2013). A mathematical model of kinetoplastid mitochondrial gene scrambling advantage. *ArXiv E-Prints*, arXIV:1307:1163 [q-bio.PE]. Retrieved from the arXiv database.

Carnes, J., McDermott, S., Anupama, A., Oliver, B. G., Sather, D. N., & Stuart, K. (2017). *In vivo* cleavage specificity of *Trypanosoma brucei* editosome endonucleases. *Nucleic Acids Research*, *45*, 4667–4686. https://doi.org/10.1093/nar/gkx116

Corell, R. A., Myler, P., & Stuart, K. (1994). *Trypanosoma brucei* mitochondrial CR4 gene encodes an extensively edited mRNA with completely edited sequence only in bloodstream forms. *Molecular and Biochemical Parasitology*, *64*(1), 65–74. https://doi.org/10.1016/0166-6851(94)90135-x

David, V., Flegontov, P., Gerasimov, E., Tanifuji, G., Hashimi, H., Logacheva, M. D., … Lukeš, J. (2015). Gene loss and error-prone RNA editing in the mitochondrion of *Perkinsela*, an endosymbiotic kinetoplastid. *MBio*, *6*(6), e01498-15. https://doi.org/10.1128/mBio.01498-15

Decker, C. J., & Sollner-Webb, B. (1990). RNA editing involves indiscriminate U changes throughout precisely defined editing domains. *Cell*, *61*(6), 1001–1011. https://doi.org/10.1016/0092-8674(90)90065-m

Dixit, S., Muller-McNicoll, M., David, V., Zarnack, K., Ule, J., Hashimi, H., & Lukeš, J. (2017). Differential binding of mitochondrial transcripts by MRB8170 and MRB4160 regulates distinct editing fates of mitochondrial mRNA in trypanosomes. *MBio*, *8*(1). https://doi.org/10.1128/mBio.02288-16

Faktorova, D., Dobakova, E., Pena-Diaz, P., & Lukeš, J. (2016). From simple to supercomplex: mitochondrial genomes of euglenozoan protists. *F1000Research*, *5*(F1000 Faculty Rev), 392. https://doi.org/10.12688/f1000research.8040.1

Feagin, J. E., Abraham, J. M., & Stuart, K. (1988). Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell*, *53*(3), 413–422. https://doi.org/10.1016/0092-8674(88)90161-4

Feagin, J. E., Jasmer, D. P., & Stuart, K. (1987). Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. *Cell*, *49*(3), 337–345. https://doi.org/10.1016/0092-8674(87)90286-8

Feagin, J. E., Shaw, J. M., Simpson, L., & Stuart, K. (1988). Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(2), 539–543. https://doi.org/10.1073/pnas.85.2.539

Feagin, J. E., & Stuart, K. (1988). Developmental aspects of uridine addition within mitochondrial transcripts of *Trypanosoma brucei*. *Molecular and Cellular Biology*, *8*(3), 1259–1265. https://doi.org/10.1128/mcb.8.3.1259

Flegontov, P., Butenko, A., Firsov, S., Kraeva, N., Eliáš, M., Field, M. C., … Lukeš, J. (2016). Genome of *Leptomonas pyrrhocoris*: A high-quality reference for monoxenous trypanosomatids and new insights into evolution of Leishmania. *Scientific Reports*, *6*(1), 23704. https://doi.org/10.1038/srep23704

Flegontov, P., Votýpka, J., Skalický, T., Logacheva, M. D., Penin, A. A., Tanifuji, G., … Lukeš, J. (2013). Paratrypanosoma is a novel early-branching Trypanosomatid. *Current Biology*, *23*(18), 1787–1793. https://doi.org/10.1016/j.cub.2013.07.045

Gao, G., Kapushoc, S. T., Simpson, A. M., Thiemann, O. H., & Simpson, L. (2001). Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: An unexpected complexity. *RNA*, *7*(9), 1335–1347. https://doi.org/10.1017/s1355838201018076

Gazestani, V. H., Hampton, M., Shaw, A. K., Salavati, R., & Zimmer, S. L. (2018). Tail characteristics of *Trypanosoma brucei* mitochondrial transcripts are developmentally altered in a transcript-specific manner. *International Journal for Parasitology*, *48*(2), 179–189. https://doi.org/10.1016/j.ijpara.2017.08.012

Gerasimov, E. S., Gasparyan, A. A., Kaurov, I., Tichý, B., Logacheva, M. D., Kolesnikov, A. A., … Flegontov, P. (2018). Trypanosomatid mitochondrial RNA editing: Dramatically complex transcript repertoires revealed with a dedicated mapping tool. *Nucleic Acids Research*, *46*(2), 765–781. https://doi.org/10.1093/nar/gkx1202

Gerasimov, E. S., Kostygov, A. Y., Yan, S., & Kolesnikov, A. A. (2012). From cryptogene to gene? ND8 editing domain reduction in insect trypanosomatids. *European Journal of Protistology*, *48*(3), 185–193. https://doi.org/10.1016/j.ejop.2011.09.002

Gommans, W. M., Mullen, S. P., & Maas, S. (2009). RNA editing: A driving force for adaptive evolution? *BioEssays*, *31*(10), 1137–1145. https://doi.org/10.1002/bies.200900045

Gray, M. W. (2012). Evolutionary origin of RNA editing. *Biochemistry*, *51*(26), 5235–5242. https://doi.org/10.1021/bi300419r

Gray, M. W., Lukeš, J., Archibald, J. M., Keeling, P. J., & Doolittle, W. F. (2010). Cell biology. Irremediable complexity? *Science*, *330*(6006), 920–921. https://doi.org/10.1126/science.1198594

Greif, G., Rodriguez, M., Reyna-Bello, A., Robello, C., & Alvarez-Valin, F. (2015). Kinetoplast adaptations in American strains from *Trypanosoma vivax*. *Mutation Research: Fundamental and Molecular Mechanisms of Mutagenesis*, *773*, 69–82. https://doi.org/10.1016/j.mrfmmm.2015.01.008

Hashimi, H., Cicova, Z., Novotna, L., Wen, Y. Z., & Lukeš, J. (2009). Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. *RNA*, *15*(4), 588–599. https://doi.org/10.1261/rna.1411809

Hashimi, H., Zimmer, S. L., Ammerman, M. L., Read, L. K., & Lukeš, J. (2013). Dual core processing: MRB1 is an emerging kinetoplast RNA editing complex. *Trends in Parasitology*, *29*(2), 91–99. https://doi.org/10.1016/j.pt.2012.11.005

Horváth, A., Berry, E. A., & Maslov, D. A. (2000). Translation of the edited mRNA for cytochrome b in trypanosome mitochondria. *Science*, *287*(5458), 1639–1640. https://doi.org/10.1126/science.287.5458.1639

Horváth, A., Kingan, T. G., & Maslov, D. A. (2000). Detection of the mitochondrially encoded cytochrome c oxidase subunit I in the trypanosomatid protozoan *Leishmania tarentolae.* Evidence for translation of unedited mRNA in the kinetoplast. *Journal of Biological Chemistry*, *275*(22), 17160–17165. https://doi.org/10.1074/jbc.M907246199

Jensen, R. E., & Englund, P. T. (2012). Network news: The replication of kinetoplast DNA. *Annual Review of Microbiology*, *66*(1), 473–491. https://doi.org/10.1146/annurev-micro-092611-150057

Kafková, L., Ammerman, M. L., Faktorova, D., Fisk, J. C., Zimmer, S. L., Sobotka, R., … Hashimi, H. (2012). Functional characterization of two paralogs that are novel RNA binding proteins influencing mitochondrial transcripts of *Trypanosoma brucei*. *RNA*, *18*(10), 1846–1861. https://doi.org/10.1261/rna.033852.112

Kim, K. S., Teixeira, S. M., Kirchhoff, L. V., & Donelson, J. E. (1994). Transcription and editing of cytochrome oxidase II RNAs in *Trypanosoma cruzi*. *Journal of Biological Chemistry*, *269*(2), 1206–1211.

Kirby, L. E., & Koslowsky, D. (2017). Mitochondrial dual-coding genes in *Trypanosoma brucei*. *PLoS Neglected Tropical Diseases*, *11*(10), e0005989. https://doi.org/10.1371/journal.pntd.0005989

Kirby, L. E., Sun, Y., Judah, D., Nowak, S., & Koslowsky, D. (2016). Analysis of the *Trypanosoma brucei* EATRO 164 bloodstream guide RNA transcriptome. *PLoS Neglected Tropical Diseases*, *10*(7), e0004793. https://doi.org/10.1371/journal.pntd.0004793

Kolesnikov, A. A., Merzliak, E. M., Bessolitsyna, E. A., Fediakov, A. V., & Shoenian, G. (2003). Reduction of the edited domain of the mitochondrial A6 gene for ATPase subunit 6 in Trypanosomatidae. *Molekuliarnaia Biologiia*, *37*(4), 637–642.

Koslowsky, D., Sun, Y., Hindenach, J., Theisen, T., & Lucas, J. (2014). The insect-phase gRNA transcriptome in *Trypanosoma brucei*. *Nucleic Acids Research*, *42*(3), 1873–1886. https://doi.org/10.1093/nar/gkt973

Koslowsky, D. J., Bhat, G. J., Perrollaz, A. L., Feagin, J. E., & Stuart, K. (1990). The MURF3 gene of *T. brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. *Cell*, *62*(5), 901–911. https://doi.org/10.1016/0092-8674(90)90265-g

Koslowsky, D. J., Bhat, G. J., Read, L. K., & Stuart, K. (1991). Cycles of progressive realignment of gRNA with mRNA in RNA editing. *Cell*, *67*(3), 537–546. https://doi.org/10.1016/0092-8674(91)90528-7

Koslowsky, D. J., Riley, G. R., Feagin, J. E., & Stuart, K. (1992). Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of *Trypanosoma brucei*. *Molecular and Cellular Biology*, *12*(5), 2043–2049. https://doi.org/10.1128/mcb.12.5.2043

Kumar, V., Madina, B. R., Gulati, S., Vashisht, A. A., Kanyumbu, C., Pieters, B., … Cruz-Reyes, J. (2016). REH2C helicase and GRBC subcomplexes may base pair through mRNA and small guide RNA in kinetoplastid editosomes. *Journal of Biological Chemistry*, *291*(11), 5753–5764. https://doi.org/10.1074/jbc.M115.708164

Landweber, L. F. (1992). The evolution of RNA editing in kinetoplastid protozoa. *Bio Systems*, *28*(1–3), 41–45. https://doi.org/10.1016/0303-2647(92)90006-k

Landweber, L. F., Fiks, A. G., & Gilbert, W. (1993). The boundaries of partially edited transcripts are not conserved in kinetoplastids: Implications for the guide RNA model of editing. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(20), 9242–9246. https://doi.org/10.1073/pnas.90.20.9242

Landweber, L. F., & Gilbert, W. (1993). RNA editing as a source of genetic variation. *Nature*, *363*(6425), 179–182. https://doi.org/10.1038/363179a0

Leeder, W.-M., Hummel, N. F. C., & Göringer, H. U. (2016). Multiple G-quartet structures in pre-edited mRNAs suggest evolutionary driving force for RNA editing in trypanosomes. *Scientific Reports*, *6*(1), 29810. https://doi.org/10.1038/srep29810

Licht, K., & Jantsch, M. F. (2016). Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. *Journal of Cell Biology*, *213*(1), 15–22. https://doi.org/10.1083/jcb.201511041

Lukeš, J., Archibald, J. M., Keeling, P. J., Doolittle, W. F., & Gray, M. W. (2011). How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life*, *63*(7), 528–537. https://doi.org/10.1002/iub.489

Lukeš, J., Arts, G. J., van den Burg, J., de Haan, A., Opperdoes, F., Sloof, P., & Benne, R. (1994). Novel pattern of editing regions in mitochondrial trancripts of the cryptobiid *Trypanoplasma borreli*. *EMBO Journal*, *13*(21), 5086–5098.

Lukeš, J., Butenko, A., Hashimi, H., Maslov, D. A., Votypka, J., & Yurchenko, V. (2018). Trypanosomatids are much more than just trypanosomes. *Trends in Parasitology*, *34*(6), 466–480. https://doi.org/10.1016/j.pt.2018.03.002

Madej, M. J., Niemann, M., Hüttenhofer, A., & Göringer, H. U. (2008). Identification of novel guide RNAs from the mitochondria of *Trypanosoma brucei*. *RNA Biology*, *5*(2), 84–91. https://doi.org/10.4161/rna.5.2.6043

Madina, B. R., Kumar, V., Metz, R., Mooers, B. H. M., Bundschuh, R., & Cruz-Reyes, J. (2014). Native mitochondrial RNA-binding complexes in kinetoplastid RNA editing differ in guide RNA composition. *RNA*, *20*(7), 1142–1152. https://doi.org/10.1261/rna.044495.114

Madina, B. R., Kuppan, G., Vashisht, A. A., Liang, Y. H., Downey, K. M., Wohlschlegel, J. A., … Cruz-Reyes, J. (2011). Guide RNA biogenesis involves a novel RNase III family endoribonuclease in *Trypanosoma brucei*. *RNA*, *17*(10), 1821–1830. https://doi.org/10.1261/rna.2815911

Maslov, D. A. (2010). Complete set of mitochondrial pan-edited mRNAs in Leishmania mexicana amazonensis LV78. *Molecular and Biochemical Parasitology*, *173*(2), 107–114. https://doi.org/10.1016/j.molbiopara.2010.05.013

Maslov, D. A., Avila, H. A., Lake, J. A., & Simpson, L. (1994). Evolution of RNA editing in kinetoplastid protozoa. *Nature*, *368*(6469), 345–348. https://doi.org/10.1038/368345a0

Maslov, D. A., Hollar, L., Haghighat, P., & Nawathean, P. (1998). Demonstration of mRNA editing and localization of guide RNA genes in kinetoplast-mitochondria of the plant trypanosomatid *Phytomonas serpens*. *Molecular and Biochemical Parasitology*, *93*(2), 225–236. https://doi.org/10.1016/s0166-6851(98)00028-0

Maslov, D. A., Nawathean, P., & Scheel, J. (1999). Partial kinetoplast-mitochondrial gene organization and expression in the respiratory deficient plant trypanosomatid *Phytomonas serpens*. *Molecular and Biochemical Parasitology*, *99*(2), 207–221. https://doi.org/10.1016/s0166-6851(99)00028-6

Maslov, D. A., & Simpson, L. (1994). RNA editing and mitochondrial genomic organization in the cryptobiid kinetoplastid protozoan *Trypanoplasma borreli*. *Molecular and Cellular Biology*, *14*(12), 8174–8182. https://doi.org/10.1128/mcb.14.12.8174

Maslov, D. A., Sturm, N. R., Niner, B. M., Gruszynski, E. S., Peris, M., & Simpson, L. (1992). An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. *Molecular and Cellular Biology*, *12*(1), 56–67. https://doi.org/10.1128/mcb.12.1.56

Maslov, D. A., Votýpka, J., Yurchenko, V., & Lukeš, J. (2013). Diversity and phylogeny of insect trypanosomatids: All that is hidden shall be revealed. *Trends in Parasitology*, *29*(1), 43–52. https://doi.org/10.1016/j.pt.2012.11.001

McAdams, N. M., Simpson, R. M., Chen, R., Sun, Y., & Read, L. (2018). MRB7260 is essential for productive protein—RNA interactions within the RNA editing substrate binding complex during trypanosome RNA editing. *RNA*, *24*, 540–556. https://doi.org/10.1261/rna.065169.117

Merzlyak, E. M., Zakharova, M. Y., & Kolesnikov, A. A. (2001). Monogenetic trypanosomatids: Comparison of the ND8 editing gene. *European Journal of Protistology*, *37*(2), 233–240. https://doi.org/10.1078/0932-4739-00823

Nawathean, P., & Maslov, D. A. (2000). The absence of genes for cytochrome c oxidase and reductase subunits in maxicircle kinetoplast DNA of the respiration-deficient plant trypanosomatid *Phytomonas serpens*. *Current Genetics*, *38*(2), 95–103. https://doi.org/10.1007/s002940000135

Nebohácová, M., Kim, C. E., Simpson, L., & Maslov, D. A. (2009). RNA editing and mitochondrial activity in promastigotes and amastigotes of *Leishmania donovani*. *International Journal for Parasitology*, *39*(6), 635–644. https://doi.org/10.1016/j.ijpara.2008.10.015

Nisbet, R. E. R., Kurniawan, D. P., Bowers, H. D., & Howe, C. J. (2016). Transcripts in the Plasmodium apicoplast undergo cleavage at tRNAs and editing, and include antisense sequences. *Protist*, *167*(4), 377–388. https://doi.org/10.1016/j.protis.2016.06.003

Ochs, D. E., Otsu, K., Teixeira, S. M., Moser, D. R., & Kirchhoff, L. V. (1996). Maxicircle genomic organization and editing of an ATPase subunit 6 RNA in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, *76*(1–2), 267–278. https://doi.org/10.1016/0166-6851(95)02565-0

Ochsenreiter, T., Anderson, S., Wood, Z. A., & Hajduk, S. L. (2008). Alternative RNA editing produces a novel protein involved in mitochondrial DNA maintenance in trypanosomes. *Molecular and Cellular Biology*, *28*(18), 5595–5604. https://doi.org/10.1128/MCB.00637-08

Ochsenreiter, T., Cipriano, M., & Hajduk, S. L. (2008). Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PLoS One*, *3*(2), e1566. https://doi.org/10.1371/journal.pone.0001566

Ochsenreiter, T., & Hajduk, S. L. (2006). Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Reports*, *7*(11), 1128–1133. https://doi.org/10.1038/sj.embor.7400817

Panigrahi, A. K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R. A., Acestor, N., … Stuart, K. D. (2009). A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. *Proteomics*, *9*(2), 434–450. https://doi.org/10.1002/pmic.200800477

Panigrahi, A. K., Zíková, A., Dalley, R. A., Acestor, N., Ogata, Y., Anupama, A., … D, K. (2008). Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. *Molecular & Cellular Proteomics*, *7*(3), 534–545. https://doi.org/10.1074/mcp.M700430-MCP200

Povelones, M. L. (2014). Beyond replication: Division and segregation of mitochondrial DNA in kinetoplastids. *Molecular and Biochemical Parasitology*, *196*(1), 53–60. https://doi.org/10.1016/j.molbiopara.2014.03.008

Ramírez, C., Puerta, C., & Requena, J. M. (2011). Evidence of RNA editing in *Leishmania braziliensis* promastigotes. *Parasitology Research*, *108*(3), 731–739. https://doi.org/10.1007/s00436-010-2190-6

Read, L. K., Fish, W. R., Muthiani, A. M., & Stuart, K. (1993). Maxicircle DNA and edited mRNA sequences of closely related trypanosome species: Implications of kRNA editing for evolution of maxicircle genomes. *Nucleic Acids Research*, *21*(17), 4073–4078. https://doi.org/10.1093/nar/21.17.4073

Read, L. K., Jacob, A. N., Fish, W. R., Muthiani, A. M., & Stuart, K. (1993). Sequences of three *Trypanosoma congolense* maxicircle genes allow prediction of regions encoding transcripts that undergo extensive RNA editing. *Molecular and Biochemical Parasitology*, *60*(2), 337–341. https://doi.org/10.1016/0166-6851(93)90146-o

Read, L. K., Lukeš, J., & Hashimi, H. (2016). Trypanosome RNA editing: The complexity of getting U in and taking U out. *WIREs RNA*, *7*(1), 33–51. https://doi.org/10.1002/wrna.1313

Read, L. K., Myler, P. J., & Stuart, K. (1992). Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei*. *Journal of Biological Chemistry*, *267*(2), 1123–1128.

Read, L. K., Stankey, K. A., Fish, W. R., Muthiani, A. M., & Stuart, K. (1994). Developmental regulation of RNA editing and polyadenylation in four life cycle stages of *Trypanosoma congolense*. *Molecular and Biochemical Parasitology*, *68*(2), 297–306. https://doi.org/10.1016/0166-6851(94)90174-0

Read, L. K., Wilson, K. D., Myler, P. J., & Stuart, K. (1994). Editing of *Trypanosoma brucei* maxicircle CR5 mRNA generates variable carboxy terminal predicted protein sequences. *Nucleic Acids Research*, *22*(8), 1489–1495. https://doi.org/10.1093/nar/22.8.1489

Riley, G. R., Corell, R. A., & Stuart, K. (1994). Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. *Journal of Biological Chemistry*, *269*(8), 6101–6108.

Ruvalcaba-Trejo, L. I., & Sturm, N. R. (2011). The Trypanosoma cruzi Sylvio X10 strain maxicircle sequence: The third musketeer. *BMC Genomics*, *12*, 58. https://doi.org/10.1186/1471-2164-12-58

Schnaufer, A., Panigrahi, A. K., Panicucci, B., Igo, R. P., Jr., Wirtz, E., Salavati, R., & Stuart, K. (2001). An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science*, *291*(5511), 2159–2162. https://doi.org/10.1126/science.1058955

Shaw, A. K., Kalem, M. C., & Zimmer, S. L. (2016). Mitochondrial gene expression is responsive to starvation stress and developmental transition in *Trypanosoma cruzi*. *mSphere*, *1*(2), e00051-16. https://doi.org/10.1128/mSphere.00051-16

Shaw, J. M., Campbell, D., & Simpson, L. (1989). Internal frameshifts within the mitochondrial genes for cytochrome oxidase subunit II and maxicircle unidentified reading frame 3 of *Leishmania tarentolae* are corrected by RNA editing: Evidence for translation of the edited cytochrome oxidase subunit II mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(16), 6220–6224. https://doi.org/10.1073/pnas.86.16.6220

Shaw, J. M., Feagin, J. E., Stuart, K., & Simpson, L. (1988). Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. *Cell*, *53*(3), 401–411. https://doi.org/10.1016/0092-8674(88)90160-2

Simpson, R. M., Bruno, A. E., Bard, J. E., Buck, M. J., & Read, L. K. (2016). High-throughput sequencing of partially edited trypanosome mRNAs reveals barriers to editing progression and evidence for alternative editing. *RNA*, *22*(5), 677–695. https://doi.org/10.1261/rna.055160.115

Simpson, R. M., Bruno, A. E., Chen, R., Lott, K., Tylec, B. L., Bard, J. E., … Read, L. K. (2017). Trypanosome RNA editing mediator complex proteins have distinct functions in gRNA utilization. *Nucleic Acids Research*, *45*(13), 7965–7983. https://doi.org/10.1093/nar/gkx458

Škodová-Sveráková, I., Horváth, A., & Maslov, D. A. (2015). Identification of the mitochondrially encoded subunit 6 of F1FO ATPase in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, *201*(2), 135–138. https://doi.org/10.1016/j.molbiopara.2015.08.002

Souza, A. E., Myler, P. J., & Stuart, K. (1992). Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited and developmentally regulated and encode a putative iron–sulfur protein homologous to an NADH dehydrogenase subunit. *Molecular and Cellular Biology*, *12*(5), 2100–2107. https://doi.org/10.1128/mcb.12.5.2100

Souza, A. E., Shu, H. H., Read, L. K., Myler, P. J., & Stuart, K. D. (1993). Extensive editing of CR2 maxicircle transcripts of *Trypanosoma brucei* predicts a protein with homology to a subunit of NADH dehydrogenase. *Molecular and Cellular Biology*, *13*(11), 6832–6840. https://doi.org/10.1128/mcb.13.11.6832

Speijer, D. (2008). Evolutionary aspects of RNA editing. In H. U. Göringer (Ed.), *RNA editing Nucleic Acids and Molecular Biology* (Vol. 20, pp. 199–227). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-540-73787-2_10

Speijer, D. (2011). Does constructive neutral evolution play an important role in the origin of cellular complexity? *BioEssays*, *33*(5), 344–349. https://doi.org/10.1002/bies.201100010

Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *Journal of Molecular Evolution*, *49*(2), 169–181. https://doi.org/10.1007/pl00006540

Sturm, N. R., Maslov, D. A., Blum, B., & Simpson, L. (1992). Generation of unexpected editing patterns in *Leishmania tarentolae* mitochondrial mRNAs: Misediting produced by misguiding. *Cell*, *70*(3), 469–476. https://doi.org/10.1016/0092-8674(92)90171-8

Suematsu, T., Zhang, L., Aphasizheva, I., Monti, S., Huang, L., Wang, Q., … Aphasizhev, R. (2016). Antisense transcripts delimit exonucleolytic activity of the mitochondrial 3′ processome to generate guide RNAs. *Molecular Cell*, *61*(3), 364–378. https://doi.org/10.1016/j.molcel.2016.01.004

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., & Brennicke, A. (2013). RNA editing in plants and its evolution. *Annual Review of Genetics*, *47*(1), 335–352. https://doi.org/10.1146/annurev-genet-111212-133519

Thiemann, O. H., Maslov, D. A., & Simpson, L. (1994). Disruption of RNA editing in *Leishmania tarentolae* by the loss of minicircle-encoded guide RNA genes. *EMBO Journal*, *13*(23), 5689–5700.

Thomas, S., Martinez, L. L., Westenberger, S. J., & Sturm, N. R. (2007). A population study of the minicircles in *Trypanosoma cruzi*: Predicting guide RNAs in the absence of empirical RNA editing. *BMC Genomics*, *8*, 133. https://doi.org/10.1186/1471-2164-8-133

Valach, M., Moreira, S., Hoffmann, S., Stadler, P. F., & Burger, G. (2017). Keeping it complicated: Mitochondrial genome plasticity across diplonemids. *Scientific Reports*, *7*(1), 14166. https://doi.org/10.1038/s41598-017-14286-z

van der Spek, H., van den Burg, J., Croiset, A., van den Broek, M., Sloof, P., & Benne, R. (1988). Transcripts from the frameshifted MURF3 gene from *Crithidia fasciculata* are edited by U insertion at multiple sites. *EMBO Journal*, *7*(8), 2509–2514.

Weng, J., Aphasizheva, I., Etheridge, R. D., Huang, L., Wang, X., Falick, A. M., & Aphasizhev, R. (2008). Guide RNA-binding complex from mitochondria of trypanosomatids. *Molecular Cell*, *32*(2), 198–209. https://doi.org/10.1016/j.molcel.2008.08.023

Westenberger, S. J., Cerqueira, G. C., El-Sayed, N. M., Zingales, B., Campbell, D. A., & Sturm, N. R. (2006). *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics*, *7*, 60. https://doi.org/10.1186/1471-2164-7-60

Yablonovitch, A. L., Deng, P., Jacobson, D., & Li, J. B. (2017). The evolution and adaptation of A to I RNA editing. *PLoS Genetics*, *13*(11), e1007064. https://doi.org/10.1371/journal.pgen.1007064

Yasuhira, S., & Simpson, L. (1995). Minicircle-encoded guide RNAs from *Crithidia fasciculata*. *RNA*, *1*(6), 634–643.

Ziková, A., Verner, Z., Nenarokova, A., Michels, P. A. M., & Lukeš, J. (2017). A paradigm shift: The mitoproteomes of procyclic and bloodstream *Trypanosoma brucei* are comparably complex. *PLoS Pathogens*, *13*(12), e1006679. https://doi.org/10.1371/journal.ppat.1006679

How to cite this article: Zimmer SL, Simpson RM, Read LK. High throughput sequencing revolution reveals conserved fundamentals of U-indel editing. *WIREs RNA*. 2018;9:e1487. https://doi.org/10.1002/wrna.1487