Trends in Genetics

CellPress

Forum

RNA Editing in Mitochondria and Plastids: Weird and Widespread

Julius Lukeš,[1,2,*] Binnypreet Kaur,[1] and Dave Speijer[3,*]

Though widespread, RNA editing is rare, except in endosymbiotic organelles. A combination of higher mutation rates, relaxation of energetic constraints, and high genetic drift is found within plastids and mitochondria and is conducive for evolution and expansion of editing processes, possibly starting as repair mechanisms. To illustrate this, we present an exhaustive phylogenetic overview of editing types.

The RNA Editing Concept

RNA editing (see Glossary) describes processes whereby RNA transcripts undergo nucleotide insertions, deletions, or substitutions, usually within coding regions [1]. Consequently, RNA sequences differ from their DNA templates, allowing for transcriptome diversity. RNA editing occurs either during transcription or post-transcriptionally and involves deaminases, nucleases, ligases, and/or polymerases. It affects mRNAs, tRNAs, rRNAs, and even miRNAs, ncRNAs, and retrotransposons. When RNA editing is extensive, original DNA templates may become unrecognizable. First described in *Trypanosoma brucei* mitochondrial transcripts, limited to four added uridines (Us) [2], editing was soon detected to occur extensively [3]. **Base conversion editing** of human apolipoprotein B mRNA, which results in two different proteins from one transcript, was discovered almost in parallel [4]. Today, RNA editing is known to occur in viruses, bacteria, protists, fungi, plants, and metazoans (Figure 1).

New mass spectrometry-based sequencing technologies led to an explosion of the known number and types of RNA editing events (Tables S1 and S2 in the supplemental information online). In particular, editing of organellar RNAs was found to be extremely abundant, diverse, and of mindboggling complexity, as exemplified by mitochondrial editing in diplonemid protists [5]. Comprehensive reviews on functional aspects and evolution of editing are available for different organisms [1,3,6–8]. We will summarize the types of editing occurring in plastids and mitochondria and discuss reasons behind their emergence, diversity, and evolution. Emerging data show that even functionally similar types of editing, such as substitutions in protists, slime molds, and plants, evolved independently and use distinct protein machineries.

Different Mechanisms of RNA Editing

There are two main types of RNA editing systems, short indels (insertions or deletions of nucleotides in mRNAs) and substitutions. Only in the case of substitution via deamination are DNA sequence and corresponding mature RNA sequence still collinear (but not identical; differing in edited sites). The extent of editing ranges from a single residue to hundreds of residues throughout an RNA molecule. Locations also vary, for instance C-to-U editing occurs in nuclear transcripts (e.g., mammalian mRNAs), organellar transcripts (e.g., mRNAs of diplonemids [5]), and may even co-occur in these compartments, as in kinetoplastid protists [3]. Nowadays, the extent and variety of organellar editing (Table S1 in the supplemental information online) really stand out as compared with the nucleus (Table S2 in the supplemental information online), which contains practically all genetic information.

Glossary

Base conversion editing: the post-transcriptional change of specific nucleotide bases by deaminases.
Constructive Neutral Evolution: the hypothesis postulating that molecular mechanisms can evolve in the absence of evolutionary benefits.
Endosymbiosis: the conversion of another cell into a eukaryotic organelle following uptake. Origin of mitochondria and plastids.
Euglenids: with diplonemids and kinetoplastids, they constitute the protist group Euglenozoa.
Eukaryogenesis: the process by which eukaryotes evolved from archaea and bacteria.
Genetic drift: changes in the population frequency of gene variants due to random (non-selected) fluctuations of the population.
Open reading frame (ORF): the uninterrupted triplet protein-coding sequence from start to stop codon.
Pentatricopeptide repeat (PPR) proteins: a large family of RNA-binding proteins. The PPR is a 35-amino acid motif, combinations of which allow highly specific RNA sequence recognition.
RNA editing: the (non-splicing) co/post-transcriptional alteration of RNA sequences, such that they differ from their DNA templates.

Organellar versus Nuclear Editing of RNA

RNA editing in organelles and in the nucleus is well studied in some protists (trypanosomes, diplonemids, slime molds, heteroloboseans, and dinoflagellates) and in sponges [9], plants [8], and assorted metazoans. It occurs either post- or co-transcriptionally. In trypanosomes, post-transcriptional editing generates (much) longer transcripts from primary transcripts, while slime molds insert extra nucleotides co-transcriptionally. Nuclear editing uses hydrolytic deamination, mostly converting adenosines to inosines (A-to-I). It can control alternative splicing and translation efficiency and can alter codons. Outside of the cellular domain, editing operates as a co-transcriptional process in viruses.

RNA Editing in Mitochondria and Plastids

Extant mitochondria evolved from a single **endosymbiotic** event, but different lineage-specific developments resulted in huge diversity, as exemplified by the

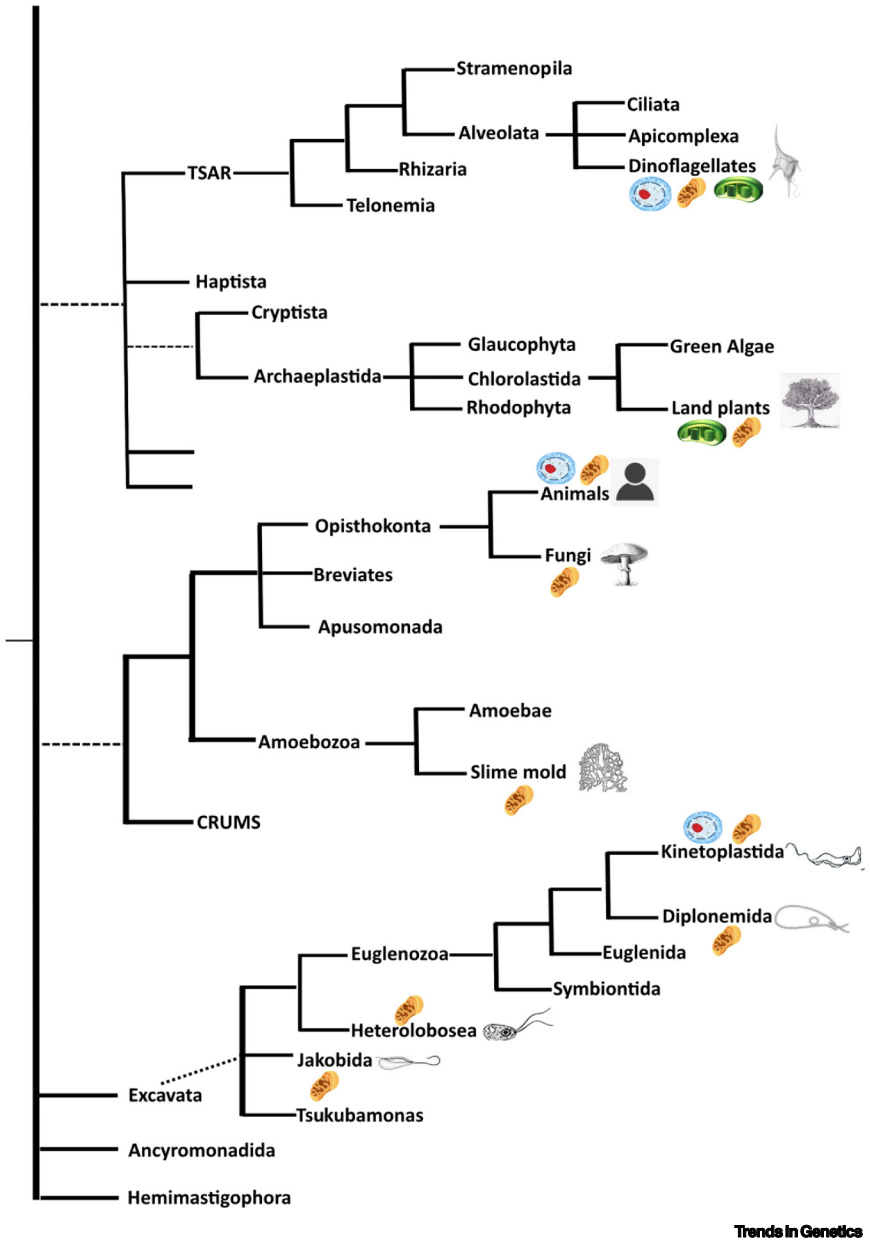

Figure 1. The Multiple Emergence of Editing across Eukaryotes. Distribution of RNA editing in the eukaryotic tree, reflecting the latest phylogenetic insights. Occurrences of editing are indicated by an image of the organism at the end of a branch. Pictograms indicate whether editing is organellar (chloroplasts and mitochondria) and/or nuclear (nuclei).

genome architecture and amount of mitochondrial DNA present in cells [10]. Processing of mitochondrial RNAs is even more varied, employing guide RNAs, *trans*-splicing and C-terminal processing, polycistronic transcription, and anti-sense RNAs, requiring dozens of imported proteins. As extreme examples, kinetoplastid and diplonemid flagellates invariably contain a single reticulated mitochondrion harboring 5–95% of total cellular DNA [11].

In trypanosomes, most protein-coding mitochondrial transcripts undergo U-insertion/deletion editing proceeding in a 3′ to 5′ direction, requiring hundreds of guide RNAs and several complexes with at least 74 imported proteins [3]. This unique, complex, extremely error-prone editing process, which involves inserting and deleting ~3000 and ~300 U residues respectively, is essential to make just a few respiratory chain subunits. Only minute fractions of transcripts are properly edited and thus become translatable. These fractions are characterized by long poly(A/U) tails that bind additional complexes, which in turn recruit mitochondrial ribosomes. Trypanosomes also use mitochondrial C-to-U substitution editing in the anticodon of a nuclear-encoded tRNA, allowing it to read UGA as tryptophan instead of stop, a widespread mitochondrion-confined departure from the genetic code. While the mitochondrial transcriptome of the related **euglenids** lacks editing [12], the sister clade of diplonemids carries huge (~250 Mbp) mitochondrial genomes consisting of circular molecules, each encoding single gene fragments [5]. Their transcripts are *trans*-spliced and edited by appending Us at fragment junctions and clustered A-to-I, C-to-U, and G-to-A substitutions. Mechanisms behind this massive splicing and editing remain unknown. Slime molds also exhibit extensive mitochondrial editing, entailing highly accurate co-transcriptional nucleotide insertions and/or deletions [7] as well as C-to-U substitutions (Table S1 in the supplemental information online).

Dinoflagellates are diverse protists that edit organellar transcripts, while their sister clades, ciliates and apicomplexans, do not, implying ‘recent’ emergence. In both plastids and mitochondria, dinoflagellates effectively perform a wide range of ‘restorative’ substitutions. In land plants, substitution editing (mostly pyrimidine transitions) is abundant in both mitochondria and plastids and is usually conserved among species [5,9]. C-to-U conversion is frequent, but reverse U-to-C switching is rare. Editing

probably arose in an ancestor that left the aquatic habitat. Did an increase in UV exposure and deleterious T-to-C mutations lead to counteracting C-to-U editing/repair? Mechanistically, organellar C-to-U editing is well understood. Specific site recognition is performed by the largest plant protein family, the **pentatricopeptide repeat (PPR) proteins,** targeted to mitochondria or (chloro)plastids. Protein-based, instead of RNA-based, sequence recognition is followed by recruitment of deaminases and endonucleases [13].

tRNAs are also subject to editing, an event quite widespread across eukaryotes [7]. Editing in mitochondrial tRNAs of the amoebas and slime molds occurs, for instance, via U-to-A/G transversions and A-to-G transitions. C-to-U tRNA editing was found in marsupial mitochondria, several plants, and kinetoplastids (see previously). tRNAs also undergo A-to-I deamination, allowing so-called anticodon wobble expansion [7].

Emergences of Editing

Editing evolved multiple times, as no functional similarity exists between U insertion/deletion editing and/or the different base conversions. Moreover, editing occurs independently in protists, plants, and animals (Figure 1) as well as in bacteria and viruses (Table S2 in the supplemental information online). Its impact also varies greatly, from virtual protein 'rebuilding' at the RNA level to correcting isolated 'errors' [fixing **open reading frames (ORFs)**, altering amino acids] and to 'invisible' changes not altering encoded proteins. Editing likely evolved from pre-existing enzymes, originally having unrelated activities. Evolutionary histories of editing mechanisms can be reconstructed; early branching plant lineages have very low frequencies, which increase in later-emerging clades. Even related diplonemids and kinetoplastids evolved fundamentally distinct U-insertion mechanisms in their mitochondria [3,5]. Thus, their mitochondrial DNA instability predates 'coping mechanisms', which fits **Constructive Neutral Evolution** scenarios [14]. Organisms 'get stuck' with errors being repaired by alternative use of pre-existing enzymatic activities, starting out humbly but leading to dedicated, highly complex systems. However, such mechanisms might have 'extra' evolutionary benefits. Framing these questions as 'Constructive Neutral Evolution versus other models' may reflect oversimplification of complex evolutionary realities. For now, we will switch from 'why?' to 'why there?'

Why Are Multiple Editing Mechanisms so Abundant in Mitochondria and Plastids?

Preponderance of editing in the endosymbiotic organelles, both in abundance and complexity (Figure 2 and Table S1 in the supplemental information online), results from several factors. First, lifestyles should be energetically forgiving of wasteful behavior (exemplified by the preponderance of unusable transcripts in trypanosome mitochondria; see previously). This is why unicellular eukaryotes predominate, with mitochondria and/or plastids providing abundant energy and the mitochondria even allowed suboptimal ATP synthesis (not so in some metazoan tissues). Second, following 'bottlenecks', random processes of **genetic drift** give rise to major, rapid changes in small populations. This characterizes populations of protists (moving into new habitats opening up after **eukaryogenesis**) and populations of organelles and/or organellar genomes inside eukaryotic cells. Thus, organelles should be exactly the hotspots of expression innovation we find them to be [10,15]. The third key factor is mutation rates. Mitochondria and plastids contain electron transport chains, consuming or producing O_2, with organellar genomes being mutated by local reactive oxygen species (ROS) production (also explaining higher editing levels in heterotrophs;

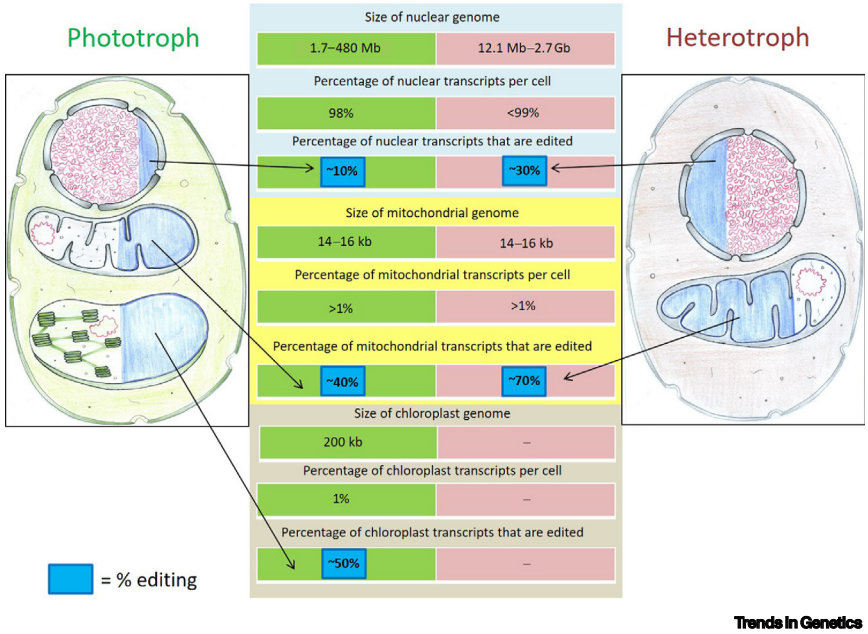

Figure 2. Typical Levels of Editing in Phototrophs and Heterotrophs. Levels of RNA editing vary extremely (both the percentage of transcripts edited and the extent of editing per transcript), as explained in the text. The general range of RNA editing in organelles and the nucleus is depicted. Overall, organelles and, in general, heterotrophs display higher levels of RNA editing, which might be related to higher amounts of damage associated with less sophisticated organellar replication mechanisms and the heterotrophic lifestyle, assuming that editing arose as a repair mechanism.

Figure 2). Did organellar editing systems start out as 'unorthodox repair mechanisms' not needed in the nucleus? ROS also contributed to the migration of organellar genes to the nucleus, where genome protection and repair improved during eukaryogenesis [15]. As to how the necessary protein machineries ended up in organelles, either enzymes without proper organellar function were inefficiently mistargeted, or *bona fide* organellar enzymes were repurposed. Whatever the mechanism(s), phylogenetic inferences show these proteins to be related to functional homologs/orthologs in the cytoplasm from which they are apparently derived. Getting rid of such 'ill-constructed' repair mechanisms is (almost) impossible, as illustrated by the preservation of extensive editing in the extremely reduced *Perkinsela* [11]; editing is here to stay.

Acknowledgements

We thank Ambar Kachale for help with the figures. This work was supported by the Czech Grant Agency 18-15962S, ERC CZ LL1601, and the ERD Funds OPVVV 0000759. We apologize to those whose references were not included due to space constraints.

Supplemental Information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tig.2020.10.004.

[1]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis), Czech Republic
[2]Faculty of Sciences, University of South Bohemia, České Budějovice (Budweis), Czech Republic
[3]Medical Biochemistry, University of Amsterdam, AMC, Amsterdam, The Netherlands

*Correspondence:
jula@paru.cas.cz (J. Lukeš) and
d.speijer@amsterdamumc.nl (D. Speijer).

https://doi.org/10.1016/j.tig.2020.10.004

References

1. Knoop, V. (2011) When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.* 68, 567–586
2. Benne, R. *et al.* (1986) Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819–8269
3. Aphasizheva, I. *et al.* (2020) Lexis and grammar of mitochondrial RNA processing in trypanosomes. *Trends Parasitol.* 36, 337–355
4. Powell, L.M. *et al.* (1987) A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* 50, 831–8404
5. Kaur, B. *et al.* (2020) Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. *Nucleic Acids Res.* 48, 2694–2708
6. Rüdinger, M. *et al.* (2011) Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. *RNA* 17, 2058–2062
7. Gott, J.M. *et al.* (2010) Two forms of RNA editing are required for tRNA maturation in *Physarum* mitochondria. *RNA* 16, 482–488
8. Sun, T. *et al.* (2016) The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.* 21, 962–973
9. Lavrov, D.V. *et al.* (2016) Extensive mitochondrial mRNA editing and unusual mitochondrial genome organization in *Calcaronean* sponges. *Curr. Biol.* 11, 86–92
10. Smith, D.R. and Keeling, P.J. (2015) Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc. Natl Acad. Sci. USA* 112, 10177–10184
11. David, V. *et al.* (2015) Gene loss and error-prone RNA editing in the mitochondrion of *Perkinsela*, an endosymbiotic kinetoplastid. *mBio* 6, 01498-15
12. Dobáková, E. *et al.* (2015) Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. *Genome Biol. Evol.* 7, 3358–3367
13. Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65, 415–442
14. Gray, M.W. *et al.* (2010) Irremediable complexity? *Science* 330, 920–921
15. Speijer, D. *et al.* (2020) Comparing early eukaryotic integration of mitochondria and chloroplasts in the light of internal ROS challenges: timing is of the essence. *mBio* 11, e00955-20

Spotlight

Noncoding RNA, Intragenomic Conflict, and Rodent SRY Evolution

Scott William Roy[1,*]

The sex-determining gene *SRY* has undergone rapid evolution in rodents. Curiously, a new study by Miyawaki *et al.* reveals that a recently evolved *SRY* gene sequence antagonizes SRY protein stability, necessitating splicing of a novel intron. Other data suggest that this troublesome gene region has noncoding RNA functions, possibly related to conflict between sex chromosomes.

In organisms with separate sexes, determination of an individual's sex is among the most important early developmental processes for reproductive fitness. Consistent with this importance, the ancient mammalian sex-determining gene *SRY* shows remarkable evolutionary conservation of mechanism and structure across a range of mammals [1]. By contrast, in rodents, *SRY* shows dynamic evolution, with various species exhibiting novel protein domains and transcripts, gene amplification, and reduced penetrance [2]. The ultimate causes of this dynamism remain unexplained

Among the most remarkable novelties of rodent *SRY* is the acquisition of novel C-terminal protein domains [3]. Recently, Miyawaki *et al.* underscored the mysterious nature of these novel C-terminal sequences. They showed that the C-terminal motif of ancestral murine SRY functions as a 'degron', a motif that causes the protein to be targeted for degradation [4]. In house mouse (*Mus musculus*), translation of the degron sequence is avoided by splicing of a novel intron sequence, producing a novel C terminus. Comparative analysis suggested that the troublesome degron-encoding sequence appeared in the ancestor of the rodent families, Muridae and Cricetidae (Figure 1A), as part of a larger region including the CAG-repeat region (called the poly-Q domain in mouse), a curious finding given that this region is thought, conversely, to stabilize the protein [3]. Consistent with the necessity of excluding the degron, various rodent lineages show diverse genomic features that prevent translation of this sequence, ranging from splicing in mouse to independent acquisition of upstream stop codons in five independent lineages to upstream frameshift mutations in three independent lineages (Figure 1B).