

Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA

R. Bundschuh^{1,2,*}, J. Altmüller³, C. Becker³, P. Nürnberg^{3,4} and J. M. Gott⁵

¹Department of Physics, ²Department of Biochemistry and Center for RNA Biology, Ohio State University, Columbus, OH 43210, USA, ³Cologne Center for Genomics, ⁴Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, D-50931 Cologne, Germany and ⁵Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH 44106, USA

Received January 31, 2010; Revised March 12, 2011; Accepted March 14, 2011

ABSTRACT

RNAs transcribed from the mitochondrial genome of *Physarum polycephalum* are heavily edited. The most prevalent editing event is the insertion of single Cs, with Us and dinucleotides also added at specific sites. The existence of insertional editing makes gene identification difficult and localization of editing sites has relied upon characterization of individual cDNAs. We have now determined the complete mitochondrial transcriptome of *Physarum* using Illumina deep sequencing of purified mitochondrial RNA. We report the first instances of A and G insertions and sites of partial and extragenic editing in *Physarum* mitochondrial RNAs, as well as an additional 772 C, U and dinucleotide insertions. The notable lack of anti-sense RNAs in our non-size selected, directional library argues strongly against an RNA-guided editing mechanism. Also of interest are our findings that sites of C to U changes are unedited at a significantly higher frequency than insertional editing sites and that substitutional editing of neighboring sites appears to be coupled. Finally, in addition to the characterization of RNAs from 17 predicted genes, our data identified nine new mitochondrial genes, four of which encode proteins that do not resemble other proteins in the database. Curiously, one of the latter mRNAs contains no editing sites.

INTRODUCTION

The production of mature RNAs often involves a complex array of events. Some transcripts contain site-specific changes that could have been encoded within the DNA

template but are not; these RNAs are said to be ‘edited’. RNA editing takes many forms, and can include alterations at either the base or nucleotide level (1). Base changes within RNAs frequently involve deamination of C to U or A to I (2,3,4), but many other types of sequence changes have been observed (1,4,5). In some cases, substitutional editing proceeds via a process involving nucleotide excision and replacement, as exemplified by the 5′ editing of mitochondrial tRNAs that contain mismatches in the acceptor stem (6,7). Changes at the nucleotide level can be quite extensive in some species, leading to the creation of functional RNAs in organelles that often lack identifiable genes. Internal addition of non-encoded nucleotides and the deletion of encoded nucleotides can proceed via multiple mechanisms, with distinct differences between species (1,8).

The mRNAs, tRNAs and rRNAs transcribed from the mitochondrial genome of *Physarum polycephalum* contain non-templated nucleotides (9,10). Nucleotide insertions are frequent events, typically making up ~4% of mRNAs and ~2% of structural RNAs. Approximately 90% of the known insertions involve single C residues, with the rest comprised of U, AA, UU, GU/UG, CU/UC, GC/CG and UA insertions. These extra nucleotides are added during transcription via a process that is mechanistically distinct from other types of insertional editing (11). Other, rarer forms of editing in *P. polycephalum* mitochondria include the deletion of three adjacent encoded A residues within the nad2 mRNA (12), four C to U changes within the cox1 mRNA (13) and two instances of nucleotide addition at the 5′-end of mitochondrially encoded tRNAs (14). The latter two forms of editing are post-transcriptional, rather than co-transcriptional events (14,15). All alterations are highly specific and very efficient; most *Physarum* mitochondrial transcripts are fully edited at all sites.

Although the complete sequence of the 62 862 bp *Physarum* mitochondrial genome is known (16), the lack

*To whom correspondence should be addressed. Tel: +1 (614) 688 3978; Fax: +1 (614) 292 7557; Email: bundschuh@mps.ohio-state.edu

of open reading frames (ORFs) corresponding to known mitochondrial genes renders standard gene-finding programs relatively ineffective. The genome does contain 20 ORFs but, curiously, their predicted polypeptide products do not resemble proteins in published databases. Development of specialized algorithms that take into account the possibility of frequent reading frame shifts resulted in the localization of a number of additional genes, some of which have been verified by cDNA sequencing (12,17). However, thus far, the sequence of only 13 mitochondrial mRNAs and 8 structural RNAs have been reported, with an additional 17 genes predicted but not characterized at the RNA level (17). In addition, a number of genes that are typically found in mitochondrial genomes have yet to be identified in *Physarum* and gaps remain in the mitochondrial gene map (17).

While RNA editing is relatively widespread across different organisms, there are very few instances in which genome-wide information on RNA editing is available. In plant chloroplasts and mitochondria, where editing is substitutional and thus does not obscure the identity of genes, complete genome-wide sets of editing sites have been determined using a painstaking gene by gene approach (18,19). More recently, it was demonstrated that the full editome of a previously not studied (namely, grape) mitochondrion can be obtained by high-throughput sequencing of mitochondrial RNA (20); the same technique also found 10 previously uncharacterized editing sites in *Arabidopsis thaliana* mitochondrial RNA (20) which was believed to be fully characterized. While RNA editing in humans is substitutional just as in plant mitochondria, the size of the human genome makes genome-wide studies of human RNA editing difficult even with high-throughput sequencing. Thus, one study (21) limits itself to a systematic characterization of a set of (tens of thousands) computationally predicted potential human editing sites, while another study uses high-throughput sequencing to determine the details of editing in one specific human transcript (22).

Here, we present the characterization of the complete set of mitochondrial transcripts synthesized by plasmodia during logarithmic growth, including the identification of the entire set of RNA editing events that occur in *Physarum* mitochondria. An extensive analysis of sequence context and codon position biases is also presented. We describe the discovery and confirmation of two new types of nucleotide insertions, further expanding the known repertoire of editing events in this organelle. Our deep-sequencing data more than double the number of characterized editing sites and mRNAs, allowing us to identify nine new mitochondrial genes and extragenic editing sites. Additional editing sites were also found within previously characterized RNAs; each of these latter sites have been confirmed by RT-PCR and/or primer extension sequencing of bulk RNA. Importantly, the depth of our sequence coverage allows us to assess the extent of editing at each site; results for both insertional and substitutional editing sites are discussed. Finally, we report the identification of a number of mitochondrial tRNAs that are encoded in the nuclear genome.

MATERIALS AND METHODS

Isolation of RNA from gradient-purified mitochondria

Plasmodial strain M3CVIII (kindly provided by Dr Mark Adelman), the strain used in most previous functional studies on *Physarum* editing, was grown as macroparasitoid at 26°C in semi-defined medium (23). Macroparasitoid were harvested directly into ice-cold BSS [10 mM Tris-HCl (pH 7.5)/0.25 M sucrose], with all subsequent steps carried out at 4°C. Cells were lysed in a Waring blender using two 15 s bursts at half maximum speed. The homogenate was filtered prior to pelleting the nuclei and remaining cell debris by centrifugation for 5 min at 700 g. Mitochondria were pelleted by centrifugation for 5 min at 5800g, resuspended in BSS, layered over 11 ml Percoll step gradients [2–3 ml layers with densities of 1.044, 1.062, 1.082 and 1.095 g/ml in 1 mM Tris-HCl (pH 7.5)/0.25 M sucrose] and centrifuged at 47 800 g for 30 s. Mitochondrial fractions were collected from each gradient, diluted slowly with 2.5 vol of 1 mM Tris-HCl (pH 7.5)/0.25 M sucrose and pelleted by centrifugation at 7600g for 5 min. Total mitochondrial RNA was isolated using TRIzol reagent (Invitrogen) as specified by the manufacturer. Residual DNA was removed by digestion with DNaseI (Roche) in the supplied buffer.

Library preparation and sequencing

To preserve strand information when sequencing the RNA, we used the Illumina protocol 'Directional mRNA Library Preparation, Pre-Release Protocol Rev.A.', adapting it for mtRNA sequencing. Because mitochondrial RNAs from *Physarum* generally lack a polyA tail, the polyA-selection step was omitted and total mitochondrial RNA was fragmented using 15 µl (~500 ng) of mtRNA, 2 µl Frag Buffer (Illumina) and 3 µl ddH₂O. Incubation time was extended to 20 min (94°C, stopped with 1 µl Stop Buffer, Illumina). An Agilent Bioanalyzer run was conducted showing the mtRNA reduced to fragments between 25 and 100 bp. We decided not to implement any size selection to avoid losing naturally occurring small RNAs. The further steps of the protocol involved purification, phosphatase treatment, PNK treatment, again purification, adapter ligation, reverse transcription, amplification and again purification using AMPure beads. The final Agilent Bioanalyzer process control run showed a library size of 168 bp on average, composed of 73 bp adapter sequences and ~95 bp insert size in which some of these inserts were composed of more than one ligated original fragment. A 10 pM dilution of the library was loaded on one lane of a single read Illumina flowcell using the Illumina Cluster Station and sequenced in a 1 × 51 bp single read run on the Illumina GAIIx, following the standard protocols.

Read analysis

In the 12 831 431 raw reads all bases with quality scores below 20 were replaced by Ns. Reads that contained at least the first five bases of the known adapter sequence were truncated at the adapter sequence and all trailing Ns were removed. All reads that after this truncation

comprised at least 15 bases and no more than 3 Ns were accepted as high quality reads for further processing.

Identification of editing sites

The high-quality reads were globally aligned to the published mitochondrial genome of *P. polycephalum* (16) using match and mismatch scores of 1 and -1 and a linear gap cost of 1, respectively. Reads with a global alignment score of at least 35 and no more than three substitutions and five insertions or deletions were considered mapped to the genome and sorted by genomic region. For each genomic region and each read mapped into this region, a local alignment with the same scoring parameters was calculated. Based on these local alignments it was counted how often which base aligned to each genomic position and how often an insertion of which base(s) occurred at this genomic position. The RNA sequences were constructed by using at each genomic position the nucleotide occurring in the majority of the reads mapped to this position followed by the inserted nucleotide(s) occurring in the reads mapped to the position if the majority of reads shows such an insertion.

In order to look specifically for partial editing sites, more stringent local alignments of the reads to the genome with a linear gap cost of three (and the same match and mismatch scores as above) were calculated. Local alignments matching editing sites within their first or last three positions were discarded in order to retain only reads that fully cover a given editing site. Among the remaining reads, the fraction of reads supporting editing was determined and a putative partial editing site was reported when this fraction fell between 20% and 80% and if the edited and unedited version was supported by at least five reads each. The ensemble of reads mapping to each putative partial editing site were manually inspected to distinguish alignment artifacts from truly partially edited sites.

Identification of genes

The reconstructed RNA sequences were translated in all six frames in order to reveal the ORFs, i.e. long coding sequence uninterrupted by a stop codon. The ORFs were compared to the protein nr database using protein BLAST (24). If significant hits were found, the protein was considered identified. If no significant hits were found, the putative protein sequence was also searched against the PFAM database (25).

Statistical analysis of sequences surrounding editing sites

We follow exactly the same analysis as in ref. (12). In short, we determine background frequencies of the 4nt separately for the three codon positions from all codons in the protein-coding transcripts. Then, we separate the unambiguous editing sites by codon position of the inserted C and align the flanking sequences for all editing sites for a given codon position based on the position of the inserted C. For each position relative to the inserted C, we calculate the relative entropy as a measure of the difference between the nucleotide

distribution observed in the flanking sequences of editing sites and the appropriate background distribution (e.g. for editing sites at the third codon position, the background distribution for the -1 position is the one for the second codon position and the background distribution for the +1 position is the one for the first codon position). In order to assign a statistical significance to the observed values of the relative entropy, we computationally generate sets of flanking sequences from the background distribution, calculate the relative entropy for these sets and record how often the relative entropy of the randomly generated sets exceeds the relative entropy of the observed flanking sequences. The analysis of the unambiguous editing sites in the non-coding RNAs proceeds identically except for the fact that separation by codon position is not necessary for either the editing sites or background frequencies.

Experimental verification of new editing types

Primer extension sequencing. End-labeled primers (12nd2, 1nd5, 34LSU, 1rpL16) were mixed with either 2 µg total mitochondrial RNA or 1.25 µg HincII-digested mitochondrial DNA in a buffer containing 50 mM Tris-HCl (pH 8.3)/60 mM NaCl/10 mM DTT in a total volume of 9 µl. RNA/primer mixes were heated to 65°C for 3 min and DNA/primer mixes were heated to 95°C for 3 min, then put immediately into a dry ice/ethanol bath. After thawing on ice, MgOAc was added to a final concentration of 4 mM and 2 µl of each primer/template mixture was distributed into a well of a microtiter plate containing all four dNTPs (final concentration 375 µM each) + one ddNTP (40 µM final concentration) in 50 mM Tris-HCl (pH 8.3)/60 mM NaCl/10 mM DTT/6 mM MgOAc (or dNTPs + buffer as a control for primer extension stops). To each well, 1.4 U of AMV reverse transcriptase (Life Sciences) diluted in the same buffer were added and the reactions were incubated for 30 min at 48°C. Reactions were stopped by the addition of 6 µl formamide loading dye. Samples were heated at 95°C for 3 min prior to loading on an 8% acrylamide/TBE/7 M urea sequencing gel. Gels were fixed and dried and the bands were visualized using a phosphorimager.

tRNA circularization products. A quantity of 10 µg total mitochondrial RNA was heated to 90°C for 5 min, cooled on ice and then incubated overnight at 37°C in 50 mM HEPES (pH 7.5)/15 mM MgCl₂/3.3 mM DTT/10% DMSO/0.01 µg/µl BSA/80 µM ATP in the presence of 15 U of T4 RNA ligase (Promega) in 20 µl. RNAs were deproteinized and ethanol precipitated prior to cDNA synthesis using a tRNA-Lys-specific primer (cirRTlys1). Primer annealing was carried out by mixing 50 nmol of primer with 2 µg of circularized RNA, heating to 90°C for 2 min and gradually cooling to room temperature. The primer/template mix was then incubated for 45 min at 42°C in 50 mM Tris-HCl (pH 7.5)/50 mM KCl/10 mM MgCl₂/10 mM DTT/0.5 mM spermidine/60 µM each dNTP in the presence of 10 U of AMV reverse transcriptase (Life Sciences) in a total vol of 30 µl. A quantity of 5 µl of the cDNA product was used in a 50 µl PCR reaction

using Taq polymerase (NEB) and kinased primers, cirRTlys1 and cirlys2 under conditions recommended by the manufacturer. The resulting PCR product was cloned into the SmaI site of pBSM13+ (Stratagene) for Sanger sequencing (Biotic Solutions).

RT-PCR. Primers for reverse transcription were mixed with 1 µg total mitochondrial RNA in 10 mM Tris-HCl (pH 8.3)/250 mM KCl, heated to 95°C for 2 min, placed at 65°C for 10 min, then put on ice. The primer/template mix was then incubated for 1 h at 42°C in 24 mM Tris-HCl (pH 8.3)/125 mM KCl/16 mM MgCl₂/8 mM DTT/400 µM each dNTP in the presence of 5 U of AMV reverse transcriptase (Life Sciences) in a total volume of 10 µl. Reactions were stopped by heating for 5 min at 95°C, followed by the addition of 4 vol of water. PCR reactions were carried out in a final vol of 50 µl, using Taq polymerase (NEB) under conditions recommended by the manufacturer with 100 ng mitochondrial DNA or 5 µl of cDNA as template; primer sets are listed in the Supplementary Data. Gel-purified PCR and RT-PCR products were sequenced directly by Biotic Solutions.

Identification of nuclear encoded RNAs

We collected all reads that could not be mapped to the mitochondrial genome and identified those that could be mapped to the draft nuclear genome of *P. polycephalum* using BLAST. The draft nuclear genome of *P. polycephalum* was produced by The Genome Center at Washington University School of Medicine in St. Louis and can be obtained from ftp://genome.wustl.edu/pub/organism/Other_Single_Celled_Organisms/Physarum_polycephalum/assembly/. We used default parameters and accepted any read with BLAST E-value below 10⁻⁴ as mapped. For contigs with at least 400 mapped reads, we first tried to identify their content by mapping them to the nr database using BLAST. We discarded contigs that either turned out to be fragments of the mitochondrial genome (these reads typically have rather short matches to the mitochondrial genome which is why they were not identified in the rather stringent original mapping to the mitochondrial genome) or that encoded ribosomal RNA. To the remaining contigs we applied the same local alignment procedure as described above in 'Identification of editing' sites albeit using the genomic contig instead of the mitochondrial genome and using the more stringent parameter setting with the linear gap cost of three. This allowed us to determine the RNA sequence transcribed from the contig. For RNAs that looked like a tRNA, we manually confirmed the presence of the non-encoded CCA tail in the reads and mapped the sequence to the tRNA structure.

RESULTS AND DISCUSSION

High-throughput sequencing

The basis of editing specificity in *Physarum* is currently unknown. It is known that insertional editing in *Physarum* mitochondria is a co-transcriptional process and is thus mechanistically distinct from the

post-transcriptional, guide RNA-mediated insertion-deletion editing observed in trypanosomes (26). In spite of these differences, one goal of this work was to determine whether RNAs that could be used to direct editing are present in *Physarum* mitochondria. Therefore, we used a non-size selected, directional library for high throughput sequencing. Mitochondria were purified on Percoll gradients prior to RNA extraction. Total RNA was fragmented and a sequencing library was produced from the fragments using a protocol that conserves strandedness of the RNAs. The sequencing library was subjected to high-throughput sequencing on an Illumina platform which resulted in 12 831 431 reads of length 51 bases each. After removal of adapter sequences and quality filtering 8 247 364 high quality reads were retained for further analysis. Details of all steps are presented in the 'Materials and Methods' section. The raw reads have been submitted to the short read archive with accession number SRP005376.

Transcript mapping and gene identification

Mitochondrial transcripts. Out of the 8 247 364 high-quality reads, 4 139 662 could be reliably matched to the published 62 862 bp *P. polycephalum* mitochondrial genome over their entire length. Of the remaining reads, a majority consisted of chimeras resulting from ligation of short (<51 nt) RNA fragments and thus, do not align to the mitochondrial genome over their full length. All the fully matched reads that covered a given position of the mitochondrial genome were used to determine the sequence of its RNA product, including editing sites (see below). Even in the genes with the lowest read coverage, ≥10 reads covered any given position; in the ribosomal RNAs, coverage was >10 000 reads per nucleotide for most positions. Coverage at base resolution is given as a spreadsheet in the Supplementary Data.

Based on our RNA-Seq data, we assembled putative transcripts, identifying ORFs and annotating the corresponding genes using translated ORFs in BLAST searches against the nr protein database. The resulting transcript map, including the genes identified within these transcripts, is shown in Figure 1. Its most striking features are the large number of potentially polycistronic transcripts, the presence of overlapping genes and the long stretches of mitochondrial DNA that are not transcribed under these growth conditions. Genes are extremely dense within the transcribed regions, with many examples of genes that partially overlap. The most extreme example involves nadG (stop codon at genomic position 20314) and rps2 (start codon at genomic position 20278), which share 39 bp.

Since the shortness of the Illumina reads (51 nt) does not allow us to unequivocally assign transcript ends, we are not able to distinguish co-transcribed genes from genes on overlapping transcripts based solely on our RNA-Seq data. However, at least a subset of these genes are co-transcribed. For example, based on the sequence of RT-PCR products spanning the 23S-17S rRNA intergenic region and the 17S rRNA through tRNA-Pro, genomic region 48432-53524 is transcribed as a long

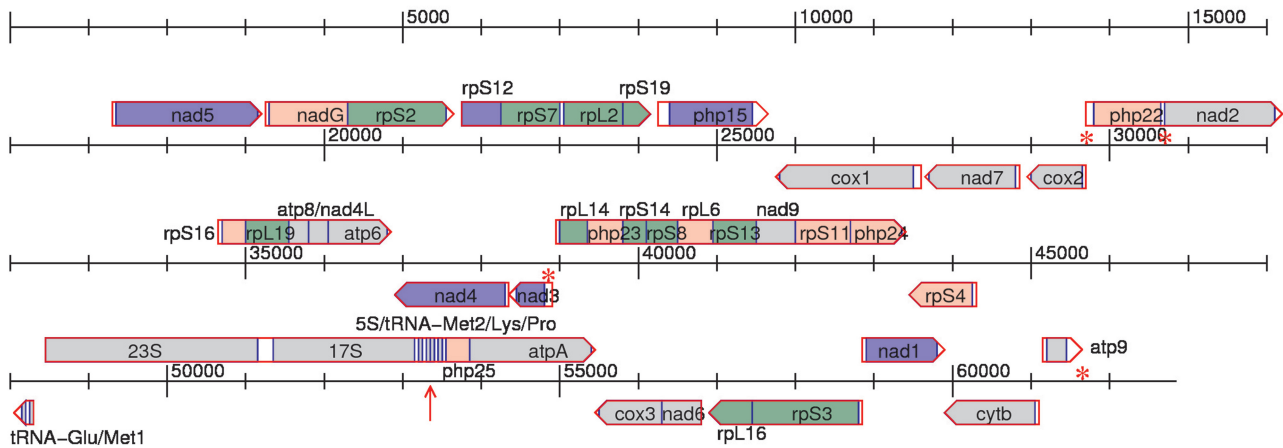


Figure 1. Transcript map of the mitochondrial genome of *P. polycephalum*. Red arrows indicate the transcripts detected in our experiment. RNAs described prior to this study are shown as gray boxes; colored boxes indicate mRNAs newly characterized in this work. Transcripts derived from newly identified genes are shaded in pink, while those from previously annotated or predicted genes are shaded in blue and green, respectively. The red asterisks show the approximate positions of extragenic C insertion sites while the red arrow indicates the position of partial editing.

precursor (14,27). Also, tRNA-Met1 and tRNA-Glu are synthesized as a longer precursor (28). Similarly, atp8/nad4L/atp6 are co-transcribed (12), but rpL19 is likely to be transcribed separately based on 5'-end mapping of the atp8 mRNA via primer extension sequencing using total mtRNA as template (data not shown). Evidence for polycistronic mitochondrial transcripts is also found in the *P. polycephalum* EST database (29) under the accession numbers EL565830 (containing parts of rpS12 and rpS7) and EL577829 (containing part of rpS13, all of nad9 and part of rpS11).

In addition to the 13 protein coding and 8 structural RNA genes for which editing sites had been determined previously, we identified 24 new edited mRNAs and 2 putative protein-coding genes whose mRNAs are not edited (see Figure 1 and Supplementary Table S1, which also includes the corresponding GenBank accession numbers). The 26 new genes include 6 genes whose approximate locations had been previously annotated in the mitochondrial genome (16), 11 predicted by Beargie *et al.* (17) and 9 genes that had escaped detection, 5 of which could be identified via BLAST searches based on our RNA-Seq data.

The four new ORFs whose identity could not be determined by BLAST searches as well as the previously annotated ORF php 15 were subjected to a PFAM-search (25). The unedited php25 mRNA has a hit to the 'mitochondrial ATP synthase B chain precursor' PFAM family (*E*-value = 0.00057). Thus, we provisionally annotate it as atpB despite its short size (97 amino acids). The N-terminal domain of php22 yields a hit to the 'rpL11, N-terminal domain' PFAM family with an *E*-value = 0.042. While this is not a statistically significant *E*-value *per se*, this weak PFAM hit suggests that php22 is a plausible (albeit highly divergent) rpL11 candidate, consistent with the presence of an rpL11 gene in the mitochondrial genomes of other amoebozoans (Gray, M., personal communication). php15, php23 and php24 do not show any notable PFAM hits.

Previously annotated ORFs

Somewhat surprisingly, a significant proportion of the *P. polycephalum* mitochondrial genome is not transcribed in plasmidia during logarithmic growth in rich medium. This does not, of course, exclude transcription under different environmental conditions or in other developmental stages. Virtually the entire 'untranscribed' region is found within the previously annotated ORFs (Supplementary Table S2). Of the 20 ORFs of unknown function in the published *P. polycephalum* mitochondrial genome, we find only ORF14 (php15) to be transcribed (and unedited) at levels comparable to the other protein coding genes. We find no evidence of insertional editing sites in any of these 'untranscribed' regions. Finally, we also note that while some strains of *Physarum* contain a mitochondrial plasmid, mF, with limited homology to the 9040–9670 region of the mitochondrial genome, we find only four scattered reads in this region in our data set and no further hits when explicitly mapping to the mF plasmid.

Our deep sequencing data are consistent with the northern analyses carried out by Jones *et al.* (30) and, with one exception, those of Takano *et al.* (16). In their initial characterization of the *Physarum* mitochondrial genome, Jones *et al.* (30) reported an 'untranscribed region' that corresponds to ORFs 1–13 and a portion of ORF20. Using PCR fragments as probes, Takano *et al.* (16) also found no evidence of transcription of ORFs 1–13 and ORF20. Band sizes observed using PCR probes to ORF14 (php15) and ORF15/16 (16) were also consistent with our results, with the ORF14 probe hybridizing to the php15 mRNA and the ORF15/16 probe detecting the php22/nad2 transcript, which overlaps their probe by ~150 nt. However, whereas the ORF17/18/19 probe detected bands of 3.7 and 4.9 kb on a northern blot (16), none of the reads in our RNA-Seq data covered any portion of the genomic region covered by their probe (45700–46965). The reasons for this discrepancy are unclear, but may be attributable to either strain differences or culture conditions.

Discrepancies with published sequences

Our deep-sequencing data diverge from previously published cDNA sequences at a limited number of positions, differing at both genomically encoded nucleotides (Supplementary Table S3) and editing sites (see below). In all eight instances of differences at genomically encoded positions, the nucleotide in our transcript agrees with the sequence of the published genome (16) while the sequences of the published cDNAs (16,31) match the genomic sequence of their respective strains. We thus conclude that these differences represent genomic variations between strains. Our edited sequences containing these variations have been submitted to GenBank with accession numbers given in Supplementary Table S1.

There is also a discrepancy between our deep-sequencing data and position 37644 of the published genome. This region is annotated as containing the *nad4* gene, but the sequence of its transcript has not been reported previously. Our reads contain an A at 37644 rather than the genomically encoded U (16). However, other *P. polycephalum* strains that have been sequenced in this region show an A in both the genomic DNA and the edited mRNA (Gott, J. and Parimi, N., unpublished data). We thus conclude that this difference is due to either a variation between strains or a genomic sequencing error.

Nuclear-encoded RNAs

Reads that could not initially be mapped to the mitochondrial genome were mapped to the draft nuclear genome of *P. polycephalum* using BLAST (24) local alignments as described in the 'Materials and Methods' section. A total of 374 201 reads mapped to a non-mitochondrial contig in this data set. A large percentage of these contigs mapped to nuclear ribosomal RNA genes. Representation of these RNAs within our library may be due to association of cytoplasmic ribosomes with the mitochondrial membrane, but this finding has not been pursued further. For the remaining nuclear contigs, RNAs transcribed from the nuclear genome and present in our mitochondrial RNA preparation were reconstructed as described in the 'Materials and methods' section. As expected, based on the fact that only five tRNAs are encoded in the mitochondrial genome, many of the remaining nuclear contigs encode tRNAs (listed in Supplementary Table S4 including the accession numbers under which they have been submitted to GenBank). Note that, this table does not include a full complement of tRNAs, which could be due to either gaps in the draft nuclear *P. polycephalum* genome or because the read coverage for a subset of tRNAs is below our chosen threshold (coverage varies considerably between tRNAs; see Supplementary Table S4).

We do not find any evidence for insertional RNA editing in the nuclearly encoded RNAs in our data set. There is, however, some disagreement between sequencing reads and genomic DNA that might be indicative of substitutional editing. However, because these discrepancies are largely localized within tRNAs at positions 26, 34 and 58, nucleotides known to be modified, these differences are

more likely a signature of reverse transcription of modified bases rather than the result of substitutional editing.

Absence of antisense RNAs

In *trypanosomatids*, editing sites are specified by guide RNAs (gRNAs) (32). The 5' portion of these small (50–75 nt) RNAs is complementary to pre-edited mRNA, forming a partial duplex that is recognized by the editing machinery, while the central region directs the insertion and/or deletion of uridines (26). Although such a mechanism seems difficult to reconcile with co-transcriptional editing, we looked for evidence of anti-sense transcripts that could conceivably be used as a form of template for nucleotide insertion in *P. polycephalum* mitochondria.

Importantly, we did not find plausible candidates for antisense RNAs that could be used to direct insertional editing within *Physarum* mitochondrial RNAs. Less than 0.01% of the reads were antisense, a level within the limits of experimental accuracy. While we cannot, of course, absolutely rule out the existence of 'guiding' RNAs, we deliberately chose conditions to maximize our chances of finding them. Considerations for library construction included: (i) use of total mitochondrial RNA, (ii) fragmentation conditions that would allow inclusion of RNAs with unusual ends (e.g. 5'-caps, 2'-3' cyclic phosphates), (iii) directional cloning and (iv) absence of size selection. Thus, our data make it extremely unlikely that antisense RNAs of any size are used to specify sites of nucleotide insertion. We note that our data do not preclude the existence of sense guide RNAs, which would be indistinguishable from reads generated by mRNA fragments. However, we have been unable to detect any guide-like RNAs using strategies that have been effective in detecting trypanosome gRNAs (33) and other small RNAs (34).

RNA editing within the mitochondrial transcriptome

Discovery of new types of editing. All previously characterized transcripts derived from the *P. polycephalum* mitochondrial genome contain one or more non-encoded nucleotides (9,12–14,16,17,31,35,36). A major goal of this work was to fully define the mitochondrial transcriptome, including the entire array of RNA editing events that occur in this organelle. Our RNA-Seq data more than double the number of known editing sites in *P. polycephalum* mitochondria, increasing the total number from 558 to 1333 (Table 1). All editing sites can be queried at <http://bioserv.mps.ohio-state.edu/redbase>. Nearly all of the editing events involve nucleotide insertion, with a total of 1347 nt added at 1324 sites. The vast majority of the newly defined editing events involve either C insertions (744 new sites) or U insertions (24 new sites). Somewhat surprisingly, only four new dinucleotide insertion sites were found, increasing their total numbers from 19 to 23; none involved novel dinucleotide combinations. It is curious that several dinucleotide combinations are not found at editing sites, although this may be just a consequence of the low total number of dinucleotide editing sites observed. No new instances of deletional or substitutional editing were found. In spite of the high frequency of

Table 1. Number of editing sites known before our study, found in our study and total number, separated by editing type

Type	Known before	This study	Total
I. Mononucleotide insertions			
C	511	744	1255
U	19	24	43
G	0	2	2
A	0	1	1
II. Dinucleotide insertions			
AA	4	0	4
UU	2	0	2
UG/GU	3	1	4
UC/CU	8	1	9
UA	1	1	2
GC/CG	1	1	2
III. Nucleotide deletions			
A deletion	3	0	3
IV. Nucleotide conversions			
U→G	1	0	1
C→G	1	0	1
C→U	4	0	4
Total	558	775	1333

Note that the sequence context of the UG/GU, UC/CU and GC/CG dinucleotide insertions makes the insertion order ambiguous. For instance, all of the UC/CU editing sites occur next to an encoded C. Thus, the resulting RNA sequence could be due to either a CU insertion (CUC) or a UC insertion (CUC). The sequence context of dinucleotide editing sites is discussed further later in the text.

editing, two protein coding genes (*php15* and *php25*) appear unedited. A similar observation has been made recently of the *nad3* gene in the related *Myxomycete Didymium iridis* (37).

Among the editing sites discovered by our approach were two single G insertions and one single A insertion, neither of which had been reported previously in *Physarum* mitochondria (Supplementary Table S5). G insertions in the *nad5* mRNA and 23S rRNA were confirmed by direct primer extension sequencing of mitochondrial RNA using reverse transcriptase and end-labeled primers (Supplementary Figure S1). The sequence of this region of the 23S rRNA is at odds with the published Genbank entry (accession number AF080601.1), which lacks the G at 50099 (as well as a C insertion at 50500). To further verify our findings, we isolated RT-PCR products spanning the A and G insertions as well as PCR products from the corresponding regions of the genome. Direct sequencing of these PCR fragments confirmed both the A in the *rpL16* mRNA and the G insertions within the *nad5* mRNA and 23S rRNAs, as well as the previously unreported C insertion at 50500 (Supplementary Figure S2). The updated 23S rRNA sequence has been submitted to GenBank with accession number HQ849399. Consistent with our findings, a fragment of the mitochondrial 23S rRNA represented in a *Physarum* EST library (accession number EL564349) (29) contains the inserted G at 50099 and the C insertion at 50500. Our confirmation of the A insertion at position 57109 is also consistent with the parallel discovery of an A insertion in the *rpL16* mRNA in *D. iridis*, a related *Myxomycete* (38).

While C insertions in *P. polycephalum* mitochondria appear to occur randomly throughout the transcriptome (31,39), the A and G insertions occur in specific biological contexts. The G insertion in the *nad5* mRNA and the A insertion in the *rpL16* mRNA both fall within regions that encode highly conserved amino acids, ATGGAA (met-glu) and GGAAAA (gly-lys), respectively (the inserted nucleotide is one of the underlined nucleotides—the precise location is ambiguous). Likewise, the G insertion at 50099 (editing site 38 of 23S rRNA) is needed to stabilize a conserved (40) stem in the 23S rRNA. Curiously, the added G falls opposite a CU insertion (editing sites 36/37) within the same stem (see Supplementary Figure S3). Thus, both editing events are required for the formation of this conserved element of rRNA secondary structure.

Although no internal single G insertions have been described previously, two of the tRNAs encoded in the *P. polycephalum* mitochondrial genome contain a non-encoded G at their 5'-ends (14). These are post-transcriptional processing events that are likely related to the 5'-editing mechanisms described in *Acanthamoeba* and other organisms (6,7). A significant fraction of each of these tRNAs is unedited or partially processed in *P. polycephalum* mitochondria (14). In contrast, similar to sites of co-transcriptional editing, the +G sites at 17959 and 57109 and the +A site at 57109 are fully edited *in vivo* (Supplementary Figure S2).

Discrepancies with previously published editing sites

Surprisingly, the transcripts reconstructed from our reads contain four editing sites in structural RNAs that are absent from the edited sequences deposited in GenBank. The discrepancies within the 23S rRNA (+G at 50099 and +C at 50500) have been described above. The other two newly identified editing sites are located within tRNA-Lys (Figure 2). These added Cs result in the creation of two conventional G-C base pairs within the acceptor stem, replacing a proposed G×G pair and altering the predicted tRNA 5'- and 3'-ends (28). To independently verify the existence of these C insertions and determine the termini of the mature tRNA-Lys, we circularized bulk mitochondrial tRNAs and carried out RT-PCR using primers that anneal on either side of the tRNA-Lys ligation site. Direct sequencing of the RT-PCR product confirmed the new sites of C insertion as well as the predicted tRNA ends (Supplementary Figure S4). The resulting tRNA more closely resembles the mitochondrial tRNA-Lys from *Didymium nigripes*, which contains added Cs at analogous sites within the acceptor stem (28). Our revised tRNA-Lys sequence has been submitted to GenBank with the accession number HQ849429.

Extragenic editing

Prior to our study, *P. polycephalum* mitochondrial editing sites had been known only within the coding region of mRNAs and within structural RNAs. We discovered 10 instances of C insertion in extragenic regions (indicated by the red asterisks in Figure 1 and listed in

Supplementary Table S6). The two C insertions between *php22* and *nad2* have been confirmed by primer extension sequencing of total mtRNA (Supplementary Figure S5) and those in the 5'-UTRs of *php22* and *nad3* have been confirmed via Sanger sequencing of bulk-RT-PCR products (data not shown). Four of the extragenic C insertions are within the long (~240 nt) 3'-UTR of the *atp9* mRNA, which lacks an ORF or significant BLAST hits. The final example of extragenic editing, a C insertion between tRNA-Met2 and tRNA-Lys, is described in the next section. All extragenic editing sites are annotated in GenBank with their neighboring genes.

Potential sites of partial editing

Insertional editing in *P. polycephalum* mitochondria is highly efficient *in vivo* (13). To look for evidence of partial editing, reliably matched reads were used, using cut-offs of between 20% and 80% editing (see 'Materials and Methods' section). While, nearly all editing sites are completely edited, a small number of insertion sites appeared to be edited less efficiently, as shown in Supplementary Table S7.

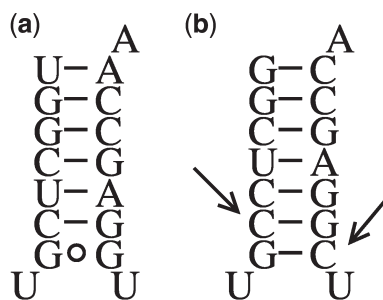


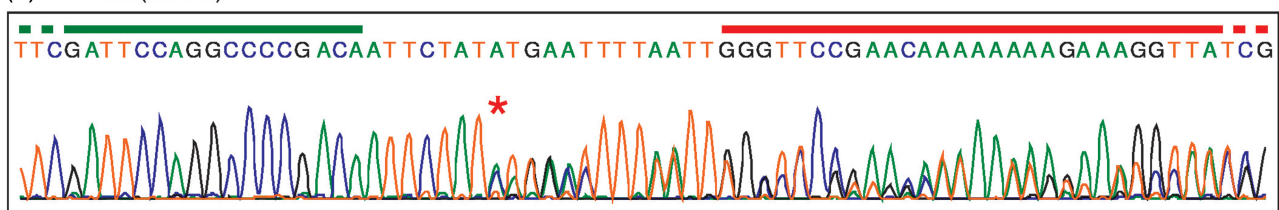
Figure 2. Acceptor stem of tRNA-Lys (a) according to (28) and (b) taking into account the two additional C insertions (arrows) and the 5'- and 3'-ends identified in this study.

To determine whether these sites are partially edited *in vivo*, we carried out RT-PCR using primers that bracket each site and sequenced the resulting PCR products directly. Sequence traces of bulk RT-PCR products covering the first four of these sites showed no evidence of partial editing (see Supplementary Figure S6), indicating that these sites are likely to be fully edited *in vivo*. This is consistent with expectations, as within coding regions the lack of an added nucleotide would shift the reading frame and lead to the production of a truncated or non-functional protein. The final site that appeared to be partially edited falls between tRNA-Met2 and tRNA-Lys, with a nearly even split between edited and unedited instances. Since this region is removed upon tRNA maturation, we used primers complementary to the RNA precursor for reverse transcription and PCR. As shown in Figure 3a the sequencing trace derived from the bulk RT-PCR product indeed doubles at the partial editing site due to the fact that roughly half of the RNAs are unedited at this site; the equivalent PCR product derived from the genome gives a uniform sequence trace throughout (Figure 3b), ruling out heteroplasmy as the source of partial editing. Thus, only at this single extragenic site, where there is likely less selective pressure to maintain efficient editing, is partial C insertion observed.

Codon bias and sequence context of C insertions

The vast increase in the number of known editing sites provided by our data prompted us to re-examine the contexts in which they are found for possible patterns. Since most types of nucleotide insertions are rare, (Table 1) this analysis was limited to C insertions, which make up 94% of all editing events. We excluded C residues inserted next to an encoded C since the exact site of insertion is ambiguous in these cases. There are 875 unambiguous C insertion sites in our data set, 797

(a) RT-PCR (mtRNA)



(b) PCR (mtDNA)

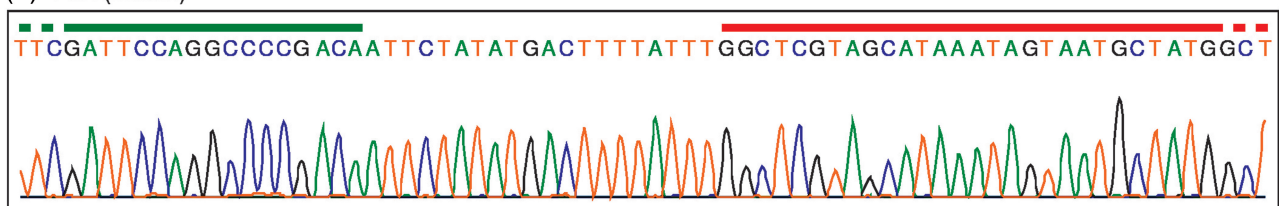


Figure 3. (a) Sequencing trace of bulk RT-PCR product encompassing the intergenic region between tRNA-Met2 and tRNA-Lys. Sequences at the 3'-end of tRNA-Met2 are indicated by a green line; the 5' portion of tRNA-Lys is indicated by a red line. Genomic position 53352 where the high-throughput sequencing reads identify a C insertion site with partial editing is indicated by a blue asterisk. Note that the sequencing trace doubles starting at this position, reflecting the mixture of edited and unedited sequences at this site. (b) Sequencing trace of bulk PCR product encompassing the same region as in (a).

of which are in protein coding genes. Of these, only 336 (269 in protein coding genes) were known before this study. One of the previously observed properties of editing sites in *Physarum* is their propensity to follow a purine–pyrimidine (31,12). We find that 59% (513 of 875) of these sites are preceded by a purine–pyrimidine, which is considerably less than the 68% (231 of 336) in our previous data set, but still significantly more than the 21% expected by chance for an arbitrary unambiguous position in the genome. Another previously observed property of *Physarum* editing sites in mRNAs is a strong bias toward the third codon position (12,31). As Supplementary Table S8 shows, in this measure, the newly characterized transcripts also show some bias, albeit much less pronounced than in the previously known mRNAs.

Supplementary Table S9 gives a more detailed view of the codon usage of all 39 mitochondrial mRNAs observed in this study. Consistent with the low GC content of the genome and the purine–pyrimidine bias discussed above, the most edited codon by far is AUC (Ile). However, every codon containing a C occurs as an edited codon at least once in the transcriptome. Also remarkable, is the observation that while the first mitochondrial mRNAs to be sequenced were all terminated by a UAA stop codon, the complete set of genes transcribed under our conditions include four ORFs terminated by UGA and one ORF terminated by UAG. The codons with the most frequent U insertions are UUA with seven and UUU with six instances.

In order to look more systematically for sequence patterns surrounding editing sites, we repeated the analysis from (12) with our much enlarged data set. The 9 nucleotides immediately upstream and downstream of the 797 unambiguous C insertions in mRNAs were extracted, based on experimental evidence that this portion of the template is required for accurate editing (41) and the 9 nt minimal spacing between inserted C's noted previously (31) and in the full set of editing sites reported here. As described in ref. (12), we separate the editing sites by codon position in order to eliminate artifacts due to the codon position bias and then determine for each codon position and each position relative to the editing site if the nucleotide distribution significantly differs from the expected distribution. In addition, we applied the same approach to the 67 unambiguous C insertions in the eight mitochondrially encoded structural RNAs. The sequence preferences for positions –6 to +6 are shown as logos in Figure 4 (none of the positions –9 to –7 or +7 to +9 showed significant deviations from background). The dashed lines in that figure indicate the limits of statistical significance based on a *P*-value of 0.05 (95% confidence interval) corrected for the $4 \times 2 \times 9 = 72$ observations. In addition to the known biases at positions –1 and –2, we find for the editing sites in coding sequences marginally significant differences at positions –3 and +1 for editing sites at the first codon position and at position +2 for editing sites at the third codon position. We do not find any significant patterns for C insertions at the second codon position beyond positions –1 and –2 due to the relatively low number of instances.

The fact that we identify different biases for different codon positions does not imply that the editing mechanism depends on codon position. It merely reflects the fact that, due to the differences in background distributions, different biases are statistically detectable for editing sites at different codon positions. In spite of the relatively small number of the editing sites in structural RNAs, several significant positions in the vicinity of the editing sites emerge. All positions show a preference for guanines, with most showing an additional preference for cytidines, perhaps reflecting selective pressure for maintenance of editing sites that fall within stable stems.

Importantly, the precise localization of editing sites cannot be accounted for based solely on the observed deviations in flanking nucleotide frequencies. Given the total length of the *P. polycephalum* mitochondrial genome ($\sim 2^{16}$), the total information content necessary to uniquely specify a single position within this genome is 16 bits. Since there are $2^2 = 4$ different nucleotides, a position within a motif in which a given nucleotide is perfectly conserved would contribute two bits of information if the GC content of the genome were 50%. For the AT-rich *Physarum* mitochondrial genome, the information content would be even higher than this for conserved G's and C's and lower for conserved A's and U's. A position in which the frequencies around the editing sites are exactly equal to the background frequencies does not contribute any information. The information content for each of the 12 positions around C insertion sites are represented by the heights of the individual bars in Figure 4. Adding up the information contained in each of the 12 positions results in only 1.9 bits for the structural RNAs and even less for protein coding genes. This is significantly less than the 16 bits required to uniquely specify a site within the full mitochondrial genome. Thus, we conclude that, although there are some unusual nucleotide frequencies around C insertion sites, they do not provide sufficient information to guide the editing machinery to these sites.

Biased contexts of non-C insertion sites

Although C insertions can occur in many contexts, sequences flanking other types of insertions are much more constrained. Whereas only 30% (380 of 1255) of C insertions occur next to an encoded C, both single Gs are inserted next to a G, the A insertion occurs next to three other As and 60% (26 of 43) of the added Us fall next to an encoded U. Thus, in a significant number of cases, the exact site of insertion cannot be determined via comparison of the RNA and genomic sequences. This bias is even more striking in the case of dinucleotide insertion sites (Supplementary Table S10). All four AA insertions occur next to an encoded A and the two UU insertions are flanked by one or two encoded Us. In the case of mixed dinucleotide insertions, even the order of the added nucleotides is usually ambiguous. This is true for all 4 UG/GU insertions, both GC/CG insertions, the 9 UC/CU insertions and one of two UA insertions (see the last column in Supplementary Table S10). We speculate that there may be an underlying mechanistic basis for

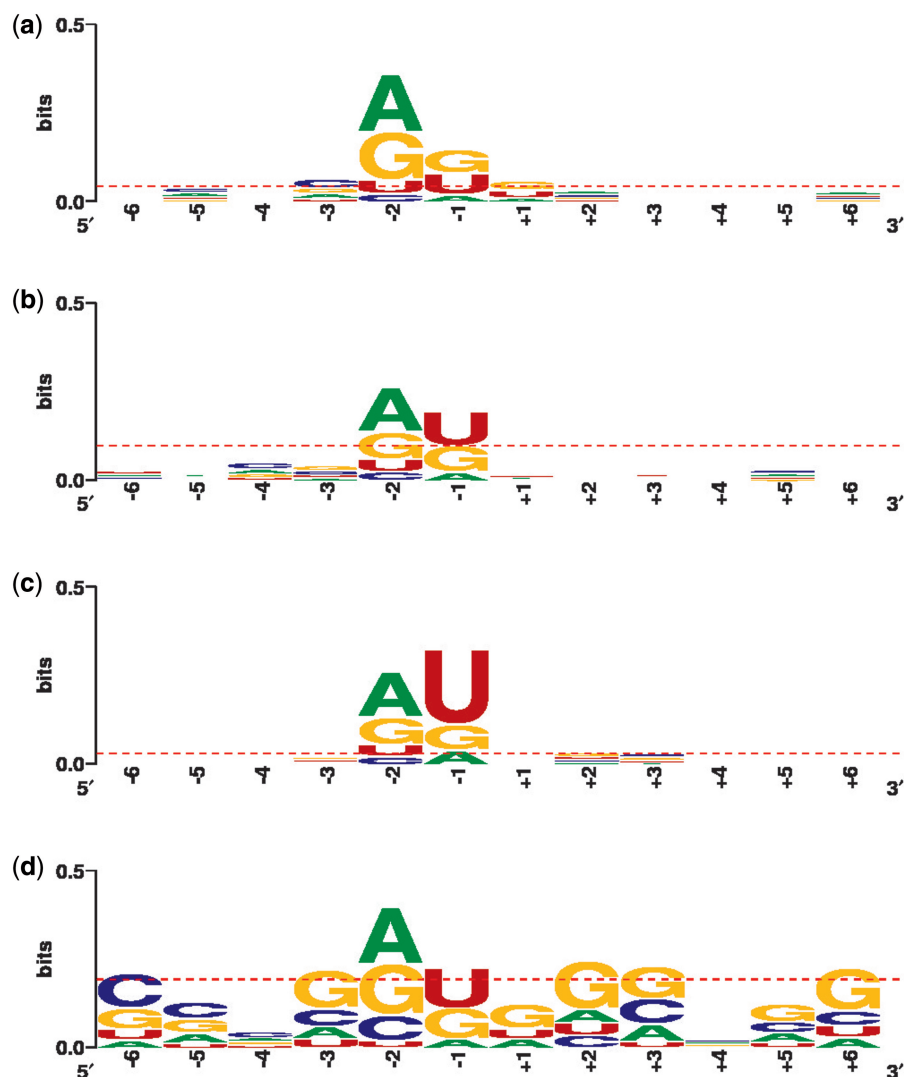


Figure 4. Deviations from expected nucleotide frequencies in the vicinity of unambiguous C insertion sites. (a), (b) and (c) correspond to editing sites in protein coding genes at the first, second and third codon position, respectively; (d) corresponds to editing sites in the structural RNAs. The relative size of the letters in one position indicates how much enhanced or suppressed the nucleotide is compared to the background distribution. The total height of the stack at a position is the total amount of information encoded by the nucleotide distribution at that position in bits. The dashed lines indicate the border of statistical significance given the total numbers of observed sites of each kind.

these ambiguities. For example, nucleotide addition at such sites could potentially be templated by polymerase stuttering at ‘slippery sites’ akin to those required for editing of paromyxoviral RNAs (42). Alternatively, templated extension after nucleotide addition might be facilitated by slippage of the RNA–DNA hybrid such that the next encoded nucleotide is added to a paired rather than an unpaired 3′-end. This feature could be particularly advantageous at dinucleotide insertion sites. This context requirement may have been lost at C insertion sites, allowing their proliferation.

Substitutional editing sites are not fully edited and show some interdependence

The *P. polycephalum* cox1 mRNA is known to have four instances of C to U editing in addition to its insertional editing sites (13). Somewhat surprisingly, we found no

evidence of additional substitutional editing elsewhere in the *P. polycephalum* mitochondrial transcriptome (Table 1). As in other organelles, C to U editing in *P. polycephalum* mitochondria is a post-transcriptional process and is thus mechanistically distinct from insertional editing (15). It was of interest, therefore, to assess the level of editing at individual substitutional editing sites in *P. polycephalum* mitochondria.

The relatively large number of reads covering the C to U sites allowed us to evaluate the extent of editing at each site. The four sites occur within two distinct regions of the cox1 mRNA; results from each group are discussed independently since our reads are too short to span the distance between the two. At the C to U site at genomic position 26779, the edited U is present in 319 of 324 reads, while only 5 reads contain the unedited C. This corresponds to 98.5% editing. Since the substitutional

Table 2. Read counts for all eight combinations of three neighboring C to U editing sites in the *cox1* mRNA

26826	26824	26823	Read count	Percentage
U	U	U	851	90.8
U	U	C	2	0.2
U	C	U	8	0.9
U	C	C	30	3.2
C	U	U	31	3.3
C	U	C	0	0.0
C	C	U	0	0.0
C	C	C	15	1.6

sequencing error rate averages $\sim 1\%$, the five reads showing the unedited C could reflect sequencing errors, leading to an underestimation of the extent of editing. However, the three encoded Us flanking this substitutional editing site are Us in all reads covering this region, giving us some confidence that the observed Cs indeed stem from unedited sequences rather than sequencing errors.

The three C to U editing sites in the second group are tightly clustered at genomic positions 26826, 26824, and 26823 (the genomic positions appear reversed since *cox1* is encoded on the reverse strand). Table 2 summarizes the read counts for all eight possible editing patterns at these three sites. Adding the percentages within the appropriate rows reveals an individual editing rate of $\sim 95\%$ for each of the three sites. However, C to U editing at these sites is clearly not independent. The two immediate neighbors at 26824 and 26823 are either both edited or both unedited 99% of the time (927/937 reads). There is also a marked correlation between the editing status of these two sites and 26826. When 26826 is unedited, 26824 and 26823 are also unedited 33% of the time (15 out of 46 instances); the expectation based on the overall rate of editing is that this would only occur in 5% of the cases. Conversely, when 26824 and 26823 are unedited, the rate at which 26826 is unedited increases from the expected 5% to 33% (15 out of 45 instances). These data indicate that, although editing at individual sites within this cluster is not obligatorily linked, there is a strong interdependence between sites, suggesting that the (as yet uncharacterized) enzyme responsible for the C to U changes is able to alter multiple sites upon binding to this region of the RNA.

SUMMARY

Maturation of *Physarum* mitochondrial transcripts requires a minimum of three different editing mechanisms. Insertion of non-encoded nucleotides (and potentially deletion of encoded nucleotides) occurs co-transcriptionally, with the extra nucleotide(s) added at the 3'-end of nascent transcripts (11). In contrast, C to U changes occur post-transcriptionally (15) and, although this process has yet to be characterized biochemically in *Physarum* mitochondria, likely entails deamination of the targeted C residues. Editing of the 5'-end of mitochondrial tRNAs is also post-transcriptional, but involves G addition opposite an encoded C within the

acceptor stem (14). Such a complex array of editing types is unprecedented, motivating us to characterize the edited transcriptome in this organelle.

The transcript map for *Physarum* mitochondria (Figure 1) shows some unusual features. Genes are densely packed in transcribed regions; many of the mRNAs are polycistronic, with numerous examples of overlapping genes. It is therefore curious that $\sim 40\%$ of the genome is not transcribed under normal growth conditions. These regions contain previously annotated ORFs that have no counterparts in the database. Of the annotated ORFs, only one, ORF14 (*php15*), is expressed in our experiments. Given that the other 19 ORFs are maintained in an organelle where genes containing ORFs are the exception rather than the rule, it seems likely that these encoded ORFs are expressed at some point in the *Physarum* life cycle and/or under growth conditions other than the one examined here.

This study is one of the first comprehensive studies of RNA editing in organelles and, to our knowledge, the first study of insertional editing using high-throughput sequencing. In the course of our work, we defined the entire set of RNA editing events in *Physarum* mitochondria, discovering and confirming two new types of editing as well as the first instances of extragenic and partial editing. A total of 775 new editing sites were identified, including 2 in the 23S rRNA and 2 in tRNA-Lys that were missed in previous experiments. The depth of our sequence coverage also provided information regarding the extent of editing at both insertion and C to U sites. Only two transcripts were not edited, *php15* and the newly identified *php25* (*atpB*) mRNA, which was not annotated as an ORF previously due to its short length. Statistical analyses of flanking nucleotides indicated that sequence context alone is not sufficient to define C insertion sites. These findings, coupled with the absence of antisense RNAs that could be used to direct editing, will significantly impact ongoing investigations into editing signals and mechanisms.

ACCESSION NUMBER

HQ849399-HQ849451.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs Michael Gray, Juan Alfonzo and Dennis Miller for useful discussions and Dr Mark Adelman for providing the *Physarum* M3C strain. R.B. thanks the Sonderforschungsbereich 680 for their hospitality during the visit in which this work was initiated.

FUNDING

National Science Foundation grants (DMR-0706002 to R.B. and SBE-0245054 to J.M.G.); the National

Institutes of Health grant (GM54663 to J.M.G.). Funding for open access charge: The National Science Foundation grant (DMR-0706002).

Conflict of interest statement. None declared.

REFERENCES

- Knoop, V. (2011) When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol. Life Sci.*, **68**, 567–586.
- Blanc, V. and Davidson, N.O. (2003) C-to-U RNA editing: mechanisms leading to genetic diversity. *J. Biol. Chem.*, **278**, 1395–1398.
- Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, **79**, 321–349.
- Chateigner-Boutin, A.L. and Small, I. (2010) Plant RNA editing. *RNA Biol.*, **7**, 213–219.
- Lin, S., Zhang, H., Spencer, D.F., Norman, J.E. and Gray, M.W. (2002) Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *J. Mol. Biol.*, **320**, 727–739.
- Price, D.H. and Gray, M.W. (1999) A novel nucleotide incorporation activity implicated in the editing of mitochondrial transfer RNAs in *Acanthamoeba castellanii*. *RNA*, **5**, 302–317.
- Bullerwell, C.E. and Gray, M.W. (2005) In vitro characterization of a tRNA editing activity in the mitochondria of *Spizellomyces punctatus*, a Chytridiomycete fungus. *J. Biol. Chem.*, **280**, 2463–2470.
- Gott, J.M. and Emeson, R.B. (2000) Functions and mechanisms of RNA editing. *Annu. Rev. Genet.*, **34**, 499–531.
- Mahendran, R., Spottswood, M.R. and Miller, D.L. (1991) RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. *Nature*, **349**, 434–438.
- Gott, J.M. and Rhee, A.C. (2007) Insertion/deletion editing in *Physarum polycephalum*. In Goringer, H.U. (ed.), *RNA editing*, Vol. 20. Springer, Berlin, pp. 85–104.
- Cheng, Y.W., Visomirski-Robic, L.M. and Gott, J.M. (2001) Non-templated addition of nucleotides to the 3' end of nascent RNA during RNA editing in *Physarum*. *EMBO J.*, **20**, 1405–1414.
- Gott, J.M., Parimi, N. and Bundschuh, R. (2005) Discovery of new genes and deletion editing in *Physarum* mitochondria enabled by a novel algorithm for finding edited mRNAs. *Nucleic Acids Res.*, **33**, 5063–5072.
- Gott, J.M., Visomirski, L.M. and Hunter, J.L. (1993) Substitutional and insertional RNA editing of the cytochrome c oxidase subunit 1 mRNA of *Physarum polycephalum*. *J. Biol. Chem.*, **268**, 25483–25486.
- Gott, J.M., Somerlot, B.H. and Gray, M.W. (2010) Two forms of RNA editing are required for tRNA maturation in *Physarum* mitochondria. *RNA*, **16**, 482–488.
- Gott, J.M. and Visomirski-Robic, L.M. (1998) RNA editing in *Physarum* mitochondria. In Grosjean, H. and Benne, R. (eds), *Modification and Editing of RNA*. ASM Press, Washington DC, pp. 395–411.
- Takano, H., Abe, T., Sakurai, R., Moriyama, Y., Miyazawa, Y., Nozaki, H., Kawano, S., Sasaki, N. and Kuroiwa, T. (2001) The complete DNA sequence of the mitochondrial genome of *Physarum polycephalum*. *Mol. Gen. Genet.*, **264**, 539–545.
- Beargie, C., Liu, T., Corriveau, M., Lee, H.Y., Gott, J. and Bundschuh, R. (2008) Genome annotation in the presence of insertional RNA editing. *Bioinformatics*, **24**, 2571–2578.
- Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A. and Brennicke, A. (2008) The process of RNA editing in plant mitochondria. *Mitochondrion*, **8**, 35–46.
- Grewe, F., Herres, S., Viehover, P., Polsakiewicz, M., Weisshaar, B. and Knoop, V. (2010) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.*, doi:10.1093/nar/gkq1227 [Epub ahead of print, 7 December 2010].
- Picardi, E., Horner, D.S., Chiara, M., Schiavon, R., Valle, G. and Pesole, G. (2010) Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Res.*, **38**, 4755–4767.
- Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., LeProust, E., Zhang, K., Gao, Y. and Church, G.M. (2009) Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science*, **324**, 1210–1213.
- Abbas, A.I., Urban, D.J., Jensen, N.H., Farrell, M.S., Kroeze, W.K., Mieczkowski, P., Wang, Z. and Roth, B.L. (2010) Assessing serotonin receptor mRNA editing frequency by a novel ultra high-throughput sequencing method. *Nucleic Acids Res.*, **38**, e118.
- Daniel, J.W. and Baldwin, H.H. (1964) Methods of culture for plasmodial myxomycetes. In Prescott, D.M. (ed.), *Methods in cell physiology*. Academic Press, New York, NY, pp. 9–41.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K. et al. (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
- Stuart, K.D., Schnaufer, A., Ernst, N.L. and Panigrahi, A.K. (2005) Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.*, **30**, 97–105.
- Byrne, E.M. and Gott, J.M. (2004) Unexpectedly complex editing patterns at dinucleotide insertion sites in *Physarum* mitochondria. *Mol. Cell Biol.*, **24**, 7821–7828.
- Antes, T., Costandy, H., Mahendran, R., Spottswood, M. and Miller, D. (1998) Insertional editing of mitochondrial tRNAs of *Physarum polycephalum* and *Didymium nigripes*. *Mol. Cell Biol.*, **18**, 7521–7527.
- Gloeckner, G., Golderer, G., Werner-Felmayer, G., Meyer, S. and Marwan, W. (2008) A first glimpse at the transcriptome of *Physarum polycephalum*. *BMC Genomics*, **9**, 6.
- Jones, E.P., Mahendran, R., Spottswood, M.R., Yang, Y.C. and Miller, D.L. (1990) Mitochondrial DNA of *Physarum polycephalum*: physical mapping, cloning and transcription mapping. *Curr. Genet.*, **17**, 331–337.
- Miller, D., Mahendran, R., Spottswood, M., Costandy, H., Wang, S., Ling, M.L. and Yang, N. (1993) Insertional editing in mitochondria of *Physarum*. *Semin. Cell Biol.*, **4**, 261–266.
- Blum, B., Bakalara, N. and Simpson, L. (1990) A model for RNA editing in kinetoplastid mitochondria: “guide” RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell*, **60**, 189–198.
- Maslov, D.A. and Simpson, L. (2007) Strategies of kinetoplastid cryptogene discovery and analysis. *Methods Enzymol.*, **424**, 127–139.
- Bullerwell, C.E., Burger, G., Gott, J.M., Kourennaia, O., Schnare, M.N. and Gray, M.W. (2010) Abundant 5S rRNA-like transcripts encoded by the mitochondrial genome in amoebozoza. *Eukaryot. Cell*, **9**, 762–773.
- Mahendran, R., Spottswood, M.S., Ghate, A., Ling, M.L., Jeng, K. and Miller, D.L. (1994) Editing of the mitochondrial small subunit rRNA in *Physarum polycephalum*. *EMBO J.*, **13**, 232–240.
- Wang, S.S., Mahendran, R. and Miller, D.L. (1999) Editing of cytochrome b mRNA in *Physarum* mitochondria. *J. Biol. Chem.*, **274**, 2725–2731.
- Hendrickson, P.G. and Silliker, M.E. (2010) RNA editing is absent in a single mitochondrial gene of *Didymium iridis*. *Mycologia*, **102**, 1288–1294.
- Hendrickson, P.G. and Silliker, M.E. (2010) RNA editing in six mitochondrial ribosomal protein genes of *Didymium iridis*. *Curr. Genet.*, **56**, 203–213.
- Liu, T. and Bundschuh, R. (2005) A model for codon position bias in RNA editing. *Phys. Rev. Lett.*, **95**, 088101.
- Gutell, R.R., Gray, M.W. and Schnare, M.N. (1993) A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. *Nucleic Acids Res.*, **21**, 3055–3074.
- Rhee, A.C., Somerlot, B.H., Parimi, N. and Gott, J.M. (2009) Distinct roles for sequences upstream of and downstream from *Physarum* editing sites. *RNA*, **15**, 1753–1765.
- Vidal, S., Curran, J. and Kolakofsky, D. (1990) A stuttering model for paramyxovirus P mRNA editing. *EMBO J.*, **9**, 2017–2022.