Ribosomal Incorporation of Short RNAs from Mitochondrial Genome Recombination Sites in *Toxoplasma gondii*

Sabrina Tetzlaff[1], Arne Hillebrand[1], Peter Lyko[2], Susann Wicke[2], Christian Schmitz-Linneweber[1]*

[1] Molecular Genetics, Humboldt University Berlin, Philippstr.13, 10115 Berlin, Germany.

[2] Biodiversity and Evolution, Humboldt University Berlin, Philippstr.13, 10115 Berlin, Germany.

* corresponding author;

Christian Schmitz-Linneweber

Humboldt-University of Berlin
· Molecular Genetics ·
Institute of Biology, Faculty of Life Sciences
Philippstr. 13, Bldg. 22
Rhoda-Erdmann-Haus
10115 Berlin
Germanye
Phone: ++49-(0)30-2093-49700
www: https://www.molgen.hu-berlin.de/
email: christian.schmitz-linneweber@rz.hu-berlin.de

Running title: *Toxoplasma* small rRNAs at mitoDNA block borders

Abstract

The mitochondrial genomes of Apicomplexans comprise merely three protein-coding genes, alongside a set of thirty to forty genes encoding small RNAs (sRNAs), many of which exhibit homologies to rRNA from *E. coli*. The expression status and integration of these short RNAs into ribosomes remains unclear and direct evidence for active ribosomes within Apicomplexan mitochondria is still lacking. In this study, we conducted small RNA sequencing on the Apicomplexan *Toxoplasma gondii* to investigate the occurrence and function of mitochondrial sRNAs. To enhance the depth of analysis for the sRNA sequencing outcomes, we also re-sequenced the *T. gondii* mitochondrial genome using an improved organelle enrichment protocol and Nanopore sequencing. It has been established previously that the *T. gondii* genome comprises 21 sequence blocks that undergo recombination among themselves. We here demonstrate that the arrangement of these sequence blocks follows a non-random pattern characterized by a specific number of combinations. Employing this refined genome for sRNA mapping, we find that many small RNAs originated from the junction sites between protein-coding blocks and rRNA sequence blocks. Surprisingly, such block border sRNAs were incorporated into polysomes together with canonical rRNA fragments. In conclusion, Apicomplexan ribosomes are active within polysomes and are indeed assembled through the integration of sRNAs, including novel sRNAs with merged mRNA-rRNA sequences. Our findings lead to the hypothesis that *T. gondii's* block-based genome organization enables the dual utilization of mitochondrial sequences as both messenger RNAs and ribosomal RNAs, potentially establishing a link between the regulation of rRNA and mRNA expression.

Introduction

Apicomplexan parasites are a group of intracellular protozoan pathogens that can cause infectious diseases, including malaria and toxoplasmosis. Toxoplasmosis is caused by *Toxoplasma gondii,* one of the most widespread human parasites, with a seroprevalence of more than one-quarter of humans worldwide (Molan et al. 2019). For many experimental

questions, *T. gondii* remains an ideal model system for studying the molecular biology of Apicomplexa (Szabo and Finney 2017). Given the economic and clinical impact of *T. gondii* and its apicomplexan relatives, there is continuous interest in distinctive biochemical and cellular features as potential targets for therapeutic intervention.

An idiosyncratic feature of Apicomplexa is a highly derived mitochondrial genome. apicomplexan cells possess a single mitochondrion that is respirationally active and essential for parasite survival (Melo et al. 2000; Vercesi et al. 1998; MacRae et al. 2012; Seeber et al. 1998). Nuclear genes encoding proteins with functions in mitochondria contribute strongly to *P. falciparum* and *T. gondii* cell fitness (Sidik et al. 2016; Bushell et al. 2017). Dozens of apicomplexan mitochondrial genomes have been sequenced (Berná et al. 2021b). These sequences showcase the extreme reductive evolution in apicomplexan mitochondria, setting records for the smallest mitochondrial genomes known to date, ranging in length from 6 to 11 kb (Hikosaka et al. 2013; Oborník and Lukeš 2015). Mitochondrial genome organization is evolutionarily very variable, with some species having a linear, monomeric genome (*Babesia* spp.), while others have concatenated arrays of genomes (Hikosaka et al. 2013). The *T. gondii* mitochondrial genome is composed of repetitive sequence blocks that are organized on multiple DNA molecules of varying lengths and combinations (Namasivayam et al. 2021; Berná et al. 2021a). Apicomplexan mitogenomes usually code only for three proteins (*cob, coxI, coxIII*) and have highly fragmented genes for rRNAs (Feagin et al. 2012; Seeber et al. 2020). For example, in *P. falciparum*, 34 genes for rRNA fragments are scattered across the mitochondrial DNA on both strands of the genome without any linear representation of the full-length large or small subunit rRNA (Feagin et al. 2012). How rRNA fragments are assembled into functional ribosomes remains an enigma.

Since the electron transport chain in mitochondria is essential for the survival of apicomplexans, the expression of mitochondrial genes is likely to be essential as well. In fact, a nuclear-encoded RNA polymerase targeted to mitochondria has been shown to be essential in the blood stages of *Plasmodium spp.* (Suplick et al. 1990; Ke et al. 2012). Two

mitochondrially localized RNA binding proteins from the RAP (RNA-binding domain abundant in apicomplexans) family are also essential for the survival of *P. falciparum*, although their role in RNA metabolism has not yet been determined (Hollin et al. 2022). In addition, the depletion of nuclear encoded mito-ribosomal proteins led to defects in the assembly of ETC complexes and in parasite proliferation, suggesting that mito translation is important for parasite survival (Lacombe et al. 2019; Ke et al. 2018; Shikha et al. 2022). Resistance to the antimalarial drug atovaquone in *P. falciparum* and *T. gondii* has been linked to mutations in the *cob* (*cytB*) gene of the mitochondria, further supporting the idea of active, essential translation in apicomplexan mitochondria.

Despite the clear importance of mitogenome expression in Apicomplexa, we lack a comprehensive catalog of transcripts and processing events. Genome-length, polycistronic transcripts are produced in *Plasmodium* that encompass all three open reading frames (Ji et al. 1996). The existence of full-length mRNAs was further confirmed by cDNA amplifications (Namasivayam et al. 2021) and long read sequencing of mRNAs in both *Plasmodium* and *Toxoplasma* (Lee et al. 2021; Namasivayam et al. 2021). There is no evidence of dedicated transcription initiation for individual protein-coding genes, and it is assumed that mRNAs are processed from the polycistronic precursor (Suplick et al. 1990; Rehkopf et al. 2000; Hillebrand et al. 2018; Feagin et al. 2012), but the mechanism of processing and the required machinery are unknown. Similarly, the production of rRNA fragments is achieved post-transcriptionally by processing long precursor RNAs (Ji et al. 1996). A catalog of small RNAs, mostly rRNA fragments, was described for *P. falciparum* based on small RNA sequencing (Hillebrand et al. 2018), but in other apicomplexans, including *T. gondii*, our understanding of rRNA species is limited and mostly based on predictions through sequence comparisons with *Plasmodium* (Feagin et al. 2012; Namasivayam et al. 2021).

Here, we present a novel method for enriching organellar nucleic acids in *T. gondii*. This method enabled us to investigate the structure of the mitochondrial genome through long-read sequencing at an unprecedented depth. In parallel, we conducted high-throughput sequencing

of small RNAs (<150 nt) and demonstrated that they are incorporated into polysomes. The combination of DNA sequencing results and transcriptome analysis also allowed us to identify novel transcripts, many of which originate from block boundaries and contain mixed origins from coding and noncoding regions.

Results

T. gondii organellar nucleic acids are enriched by a combination of selective membrane disruption and degradation of nucleo-cytosolic DNA and RNA

The *T. gondii* nuclear genome contains many insertions of mitochondrial DNA sequences (Ossorio et al. 1991; Gjerde 2013). To distinguish between NUMTs (nuclear DNA sequences that originated from mitochondria) and true mitochondrial sequences, it is necessary to enrich mitochondrial DNA. We modified a cell fractionation method that takes advantage of the differential cholesterol content in plasma membranes and organellar membranes (Subczynski et al. 2017; Esseiva et al. 2004). We incubated cells with the detergent digitonin, which selectively permeabilizes low-cholesterol membranes, leaving mitochondria and other organelles intact. We used a *T. gondii* strain that constitutively expresses GFP localized to the mitochondrial matrix to track the mitochondria during the purification process (Fig. 1A). After permeabilizing the plasma membrane with digitonin, we treated the cells with benzonase to degrade nucleo-cytosolic nucleic acids. Following inactivation of the benzonase, we lysed the mitochondria with high concentrations of detergents to release the soluble content.

Subsequently, we observed that mitochondrial GFP remained in the pellet fraction (P in Fig. 1B) with only a minor signal in the supernatant, indicating that the mitochondria remained largely intact. In contrast, cytosolic HSP70 largely remained in the supernatant after digitonin and benzonase treatment, indicating that the method successfully reduced cytoplasmic proteins. Upon lysis of the mitochondria using Triton X-100 (“lysis” in Fig. 1B), the GFP signal was predominantly recovered in the supernatant, demonstrating efficient lysis of the mitochondria and release of soluble components. This fraction still contains HSP70, but slight

contamination with cytosol is expected. We also found that cytosolic 18S and 26S rRNA were almost completely degraded, indicating the high efficacy of the benzonase treatment (Fig. 1C). In summary, our enhanced protocol for organellar enrichment effectively depletes cytosolic protein and RNA components and enriches mitochondrial macromolecules.

To assess the integrity of mitochondrial RNA following our protocol, we utilized small RNA gel blot hybridization for two known mitochondrial rRNA fragments, SSUD and LSUF (Namasivayam et al. 2021). Our analysis of total nucleic acids in the gel revealed effective removal of major rRNA species by benzonase treatment (Fig. 1C), while smaller degradation products accumulated (Fig. 1D lanes “Benz” and “Lysis”). In contrast, we observed clear bands for both SSUD and LSUF, indicating their protection from benzonase degradation. Notably, we also detected signals for both rRNA fragments in the supernatant of the lysis fraction, confirming successful mitochondrial lysis and recovery of mitochondrial rRNA fragments.

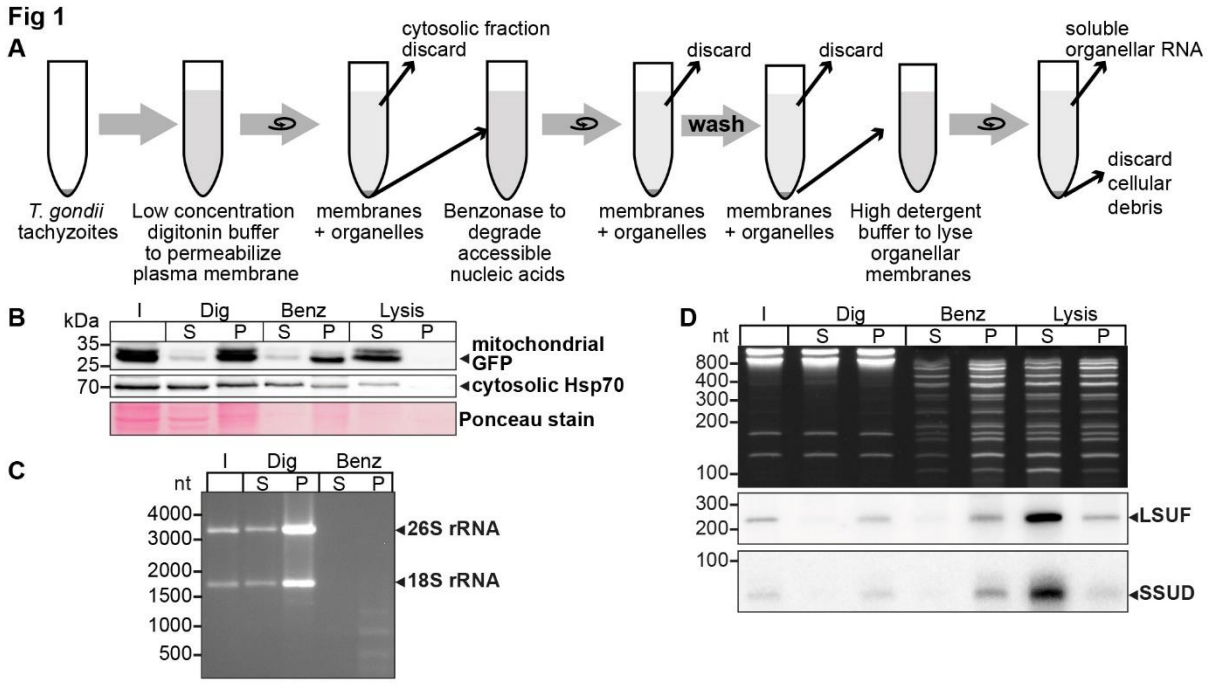

Figure 1: An experimental pipeline to enrich *T. gondii* mitochondrial RNAs.

A) *T. gondii* tachyzoites were harvested and incubated in a buffer with digitonin for plasma membrane permeabilization. Subsequently, accessible nucleic acids were digested by Benzonase. After removal of the benzonase and washing the pellet, the intact organelles are lysed by adding a detergent. Soluble nucleic acids were separated by centrifugation and extracted from the supernatant.
B) *T. gondii* cells expressing a mitochondrial-targeted GFP were subjected to organelle enrichment. GFP was tracked by immunoblotting (5% volume of each fraction was analyzed). The GFP

signal remained in the pellet and only shifted after the lysis step into the supernatant. In contrast, much of the cytosolic HSP70 is removed during the procedure, indicating specific enrichment of mitochondria.
C) RNA extracted from selected fractions of the organelle enrichment protocol was analyzed by agarose gel electrophoresis.
D) RNA extracted from selected fractions of the organelle enrichment protocol was analyzed on a denaturing 10% PAGE gel and blotted onto a nylon membrane. Radiolabeled DNA oligonucleotide probes were used to detect the mitochondrial rRNA fragments SSUD (from the small subunit of the ribosome) and LSUF (large subunit of the ribosome). Both fragments are found in pellet fractions of the protocol, where they are protected from benzonase digestion. They only shifted to the supernatant after lysis. This demonstrates that the mitochondria stay intact during the procedure.

Long-read sequencing of DNA from enriched mitochondrial fractions identifies prevalent sequence block combinations

For the analysis of mitochondrial sRNAs, we needed a reference mitochondrial genome that would represent the existing sequence block combinations. A previous sequencing study used Oxford Nanopore sequencing technology (ONT) to identify combinations of sequence blocks in *T. gondii* mitochondria (Namasivayam et al. 2021). However, a quantitative analysis of block combinations remained beyond reach due to the low amount of sequence reads. To improve read depth, we performed ONT sequencing on DNA gained from our improved organelle enrichment protocol. After filtering against the nuclear genome, we found 86,761 reads that had similarities to the published mitochondrial genome sequence (Namasivayam et al. 2021). This corresponds to 78.5 Mbp, which accounts for 9.2% of the total sequences that we attribute to the nucleus and mitochondria (Tab. S1). This is a significant increase in the sequencing depth of the *T. gondii* mitochondrial genome compared to previous attempts (Tab. S1), which can be attributed to the effectiveness of the purification process. The length of mitochondrial reads ranged from 87 nt to 17,424 nt. The GC content of these reads, on average 36.4%, supports their mitochondrial origin (Berná et al. 2021a). We annotated the reads using published sequence block information (Namasivayam et al. 2021) and confirmed the previously described 21 sequence blocks, which have been lettered from A-V (Namasivayam et al. 2021). We searched the sequencing data for repeated sequence elements but found no novel blocks, suggesting that the *T. gondii* mitochondrial genome is fully covered. We compared our reads with published ONT reads for the *T. gondii* mitochondrial genome

(Namasivayam et al. 2021; Berná et al. 2021a). While smaller reads of our dataset are found in full within longer reads in the published datasets, we do not find any examples for reads that would be full matches between the dataset. This suggests that while the block content is identical between the different sequencing results, the actual block order was reshuffled by recombination, indicating the rapidity of mitogenome evolution in *T. gondii*.

The depth of our sequencing enabled us to quantify the genome block combinations. We first asked whether there is a bias in sequential block succession. We observed that the rate of occurrence of blocks varies dramatically in our dataset, with the most frequent block, J, occurring 45,849 times and the least frequent, C, occurring 3,138 times (Fig. 2A, Tab. S2). This suggests that differences in block combinations can be expected. After counting block combinations, we found that only a small fraction of all possible block combinations are prevalent within the genome (Fig. 2B, S1A). For example, the most frequent block combination is J-B, which occurs 19,642 times in our reads. Most combinations occur less than a hundred times and make up less than 0.5% of the total number of combinations found. Fourteen blocks (R, N, M, F, L, U, I, C, K, D, H, E, T) are always found with the same preceding and following block. The direction of the blocks in these combinations is fixed (for example, the 3'-end of the L block is always fused to the 5'-end of the J block, never to the 3'-end of the J-block). This indicates that the genome's flexibility is limited and that not all block combinations are realized. Variability in genome structure is caused primarily by blocks S, O, Q, B and P, which have up to three possible neighbors, and by blocks Fp, Kp, V, A, and J, which have more than three possible neighbors. All combinations are well covered in our ONT results and helped to refine block borders relative to previous annotations (Fig. S2). Using the 31 high-frequency block combinations we found, we generated a map centered on the protein-coding genes (Fig. 2C). We will later show that all blocks not encoding proteins express rRNA fragments and are thus called rRNA blocks here. The map is not designed to represent physically existing DNA molecules, since the sequencing technology used cannot detect branched or circular DNAs, nor are the blocks drawn to scale. However, it helps to indicate the complexity of the

nonrandom recombination events shaping the genome. For example, when considering the positioning of the three protein-coding genes within this map, it is conspicuous that their three 5'-ends are always flanked by blocks also encoding a protein (Fig. 2C). This indicates an evolutionary pressure to maintain a direct connection of the 5'-terminal protein-coding sequences within the genome, with possible implications for coordinated expression of *coxI*, *coxIII*, and *cob*. In line with this, the V-Block combines the 5'-end of the *coxI* and *coxIII* coding sequences, which overlap on opposite strands (Namasivayam et al. 2021), linking these two reading frames even more tightly.

When analyzing the coding regions further, we found a large number of full representations of *cob*, *coxI* and *coxIII* in our dataset (*cob*: 1,646; *coxI*: 1,756; *coxIII*: 2,289); this accounted for 3.3%, 2.2%, and 2.3% of the total sequence length, respectively (Tab. S3). This suggests that there is no need to postulate mechanisms of genomic or posttranscriptional block shuffling to arrive at full-length open reading frames (Namasivayam et al. 2021; Berná et al. 2021b). Nevertheless, recombination within reading frames is abundant. For example, block S, as part of the *cox1* coding region, was followed in 80% of its occurrences by the rRNA block R and only in 20% by the next coding block C (Fig. 2C, Fig. S1B). Similarly, block B, as part of the *coxIII* reading frame, was followed in 66% of all cases by the rRNA block U and less frequently

by the terminal block of *coxIII*, which is block M. Whether such fusions of coding and non-coding block is of functional relevance, was unclear so far.

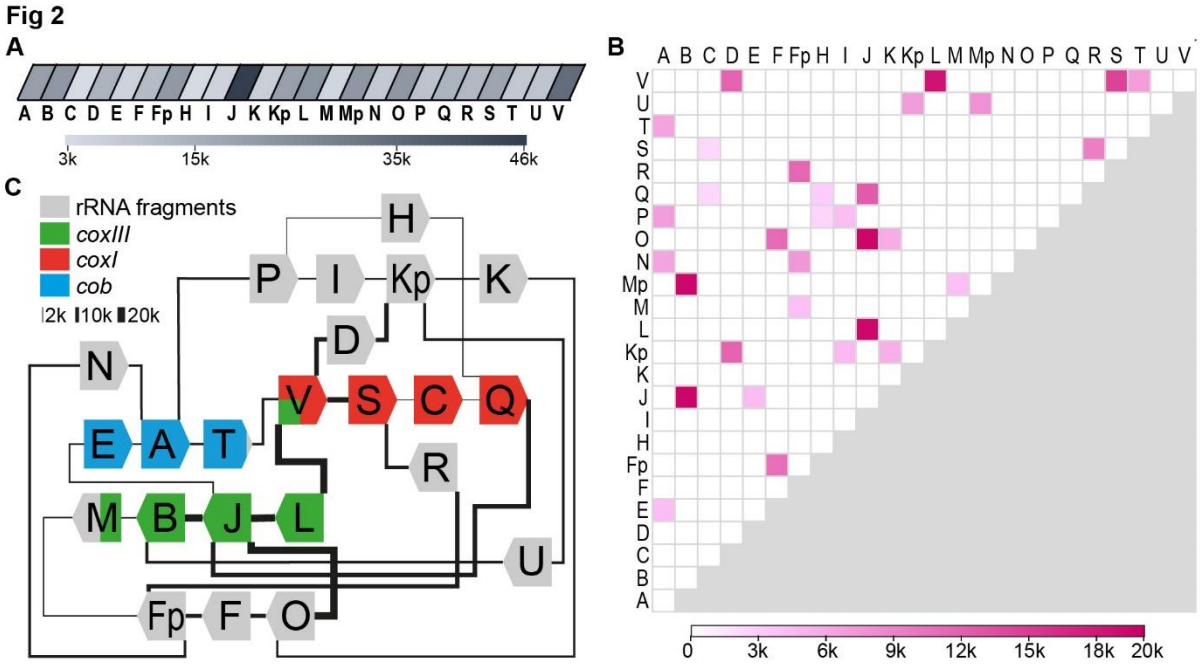

Figure 2 Long-read sequencing shows a specific system of block combinations.

ONT DNA sequencing results were analyzed for the number of blocks (A) and the number of block combinations (B). In (C), block combinations were used to generate a map of the mitochondrial genome, with the thickness of the connecting lines indicating the frequency of each combination.

Small RNA sequencing identifies a comprehensive set of mitochondrial rRNA fragments

For *T. gondii* mitochondria, rRNA species were previously predicted based on sequence comparisons to *Plasmodium* rRNA sequences (Namasivayam et al. 2021). We used Illumina-based small RNA sequencing of *T. gondii* to identify mitochondrial rRNA fragments. We sequenced both the mitochondria-enriched fraction and the input fractions described in Fig. 1. In the enriched fraction, we mostly retrieved reads for cytosolic rRNA degradation products and poor coverage of mitochondrial sequences (not shown). This is likely caused by the benzonase step in the protocol, which generates an abundance of small degradation products of cytosolic rRNA (Fig. 1D; S3). In the input samples, intact full-length rRNAs are removed by selecting for small RNAs during library preparation, resulting in less prevalent cytosolic rRNA reads. We bioinformatically removed the remaining cytosolic rRNA reads as well as any reads

shorter than 20 nt. Poly-A stretches at the 3´-ends were clipped as well. Finally, we mapped the remaining reads against the mitochondrial genome (Tab. S4) using the block combinations identified here by ONT sequencing (Fig. 2).

The results of our mapping study showed that the 5'-termini of mitochondrial small RNAs are clearly defined, making it easy to identify their transcript ends (as shown in Fig. 3A). In agreement with findings from previous studies on small RNAs from *P. falciparum* (Hillebrand et al. 2018), we found that the 3'-ends of many transcripts were more variable (as seen in the example of RNA17 in Fig. 3A). We identified a total of 34 small RNAs that accumulate in *T. gondii* mitochondria (Tab. S5). Among these, 11 correspond to previously annotated rRNA fragments (Namasivayam et al. 2021). Our sequencing data confirmed that these sequences are expressed and allowed us to refine the exact rRNA fragment ends (Tab. S5). Additionally, our small RNA sequencing revealed some larger differences compared to previous transcript predictions. This included a reassignment of SSUF to the opposite strand and affected the four rRNA fragments LSUF, LSUG, LSUD, and LSUE, which had been annotated as separate transcripts of the large subunit of the ribosome (Namasivayam et al. 2021). Our sequencing results suggest that there is accumulation of an LSUF-LSUG cotranscript and an LSUD-LSUE cotranscript. Both of these transcripts were verified via northern blotting (Fig. S4). The cotranscripts were found to be much more abundant than smaller transcript isoforms that were also detected, suggesting that the longer transcripts represent the functional rRNA fragments in *T. gondii* mitochondria. Knowing the exact sequence of rRNA fragments is crucial for further investigations into the structure of mitoribosomes, as it is key for predictions of secondary structures.

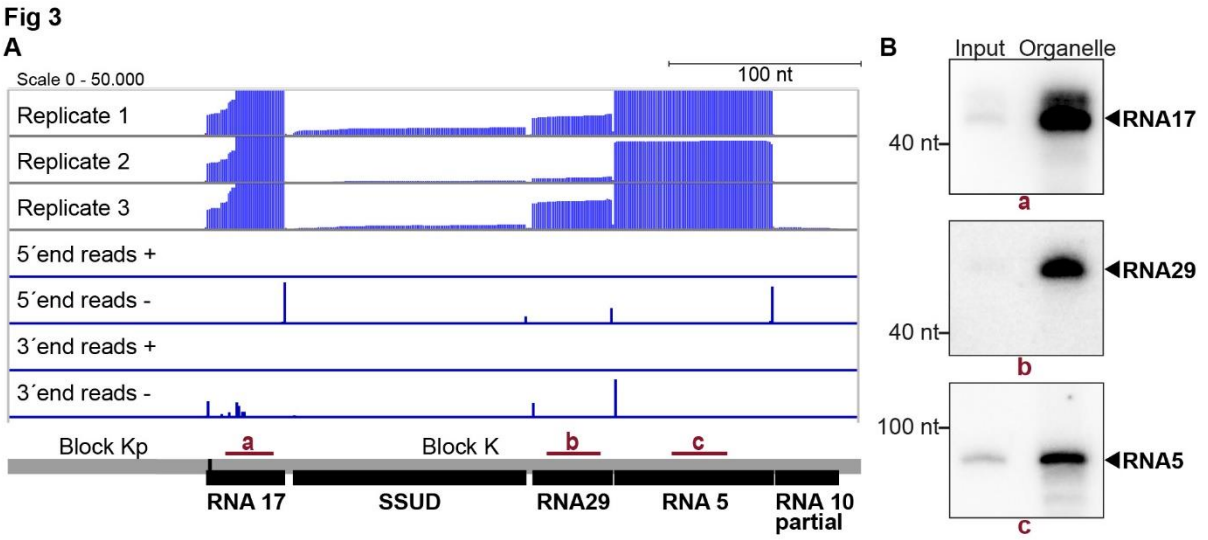

Figure 3: Small RNA sequencing identifies mitochondrial small RNAs.

(A) Upper three rows: excerpt of mapping results in three strand-specific small RNA sequencing replicates starting from total input RNA from our *T. gondii* organelle RNA preparation. Read depth was counted by the number of reads at each position. Note that the 5'-ends of the RNAs shown here are on the right. Lower four rows: only the terminal nucleotides of each read on the plus and minus strand of the genome are shown. The size and position of genes below the coverage graphs are all drawn to scale. Red bars with lowercase letters indicate positions of probes used for detecting RNAs in RNA gel blot hybridizations shown in (B).
(B) Equal quantities of total input *T. gondii* RNA and from organelle preparations were loaded onto a denaturing PAGE gel and analyzed by RNA gel blot hybridization with the probes indicated.

Discovery of novel rRNA fragments

sRNAs cover most of the noncoding sequence blocks of the T. gondii genome

Out of the 34 small RNA fragments identified, 23 have not been described previously for *T. gondii* (see Tab. S5 and Fig. S5). The 23 novel RNA fragments can be classified into two categories. First, 17 of them are homologs of rRNA fragments in the apicomplexans *P. falciparum* and *E. leuckarti*. These were named according to their *Plasmodium* homologs. The six remaining small RNA fragments are specific to *T. gondii,* and we assigned numbers extending the *Plasmodium* nomenclature (Feagin et al. 2012). Examples of novel *T. gondii* RNA species are shown in Fig. 3 for the sequence blocks Kp-K: Novel transcripts RNA5, 17, and 29, as well as the already annotated transcripts RNA10 and SSUD, are found on the minus strand (strandedness according to (Namasivayam et al. 2021; MN077088.1- MN077111.1)). They are separated by only single nucleotides in the mitochondrial genome. In this and most other noncoding blocks, the DNA sequence is almost fully used for transcript production (see

Fig. 3A, Fig. S5, S6). To validate the sequencing data, we performed RNA gel blot hybridization using probes against RNA17, RNA29, and RNA5 (Fig. 3B) and indeed detected transcripts of the expected size (42 nt and 84 nt, respectively). The transcripts were more abundant in our organelle preparations than in input samples, which indicates that they are of organellar origin and did not originate from NUMTs.

Dual use of block borders for sRNA production

Among the novel small RNAs identified, there is also a class that was only detectable due to our insights into genome block combinations. Using block combinations for mapping analysis, we identified 15 transcripts that span two blocks, i.e. there are sRNAs at almost half of all 31 block borders. Often, we find sRNAs that share a block border sequence at one end, but differ at the other end depending on the block combination (see Fig. 4A, orange boxes). An example of sequences shared at the 5'-end is represented by the pair RNA8 and RNA31, which both start in block Fp but end in different blocks (Fig. 4A). An example of block sharing at the 3'-end is the RNA pair RNA1/RNA2. Both RNAs terminate at very similar positions in block Kp but start in different blocks (Fig. 4A). Using a probe for the common 3'-part in block Kp, we detected both RNA1 and RNA2 in an RNA gel blot hybridization experiment (Fig. 4A, B). Both RNA1 and RNA2 have homology to *E. coli* LSU in their 3'-portion via block Kp, but only RNA1 has additional homology to *E. coli* LSU via block I (Fig. 4C; (Feagin et al. 2012)). It is interesting to consider that RNA1 could therefore be used in two positions in the large subunit of the ribosome. In general, block combinations lead to the expression of novel RNAs in *T. gondii* that are not found in apicomplexan species with a simpler genome organization.

Coding regions contribute to sRNAs at block borders

In addition to the three pairs of overlapping, noncoding RNAs described above (Fig. 4A), we also found five RNAs that combine sequences from coding and noncoding blocks (Fig. 5A, B). Among them, RNA16, RNA23t, and RNA34 were partially antisense to mRNAs (Fig. 5A), and

we confirmed the accumulation of RNA34 by RNA gel blot hybridization (Fig. 5C). None of the three RNAs had detectable homologies to rRNA. For all five coding-noncoding RNAs, homologs are present in the mitochondrial genome of *P. falciparum*. In *P. falciparum however,* none of them are antisense to coding regions. Whether these sRNAs form sRNA:mRNA interactions and whether this has functional consequences remains to be investigated.

Two of the small RNAs contained coding sequences in the sense direction (Fig. 5B). Both were validated by RNA gel blot hybridization (Fig. 5D). RNA19 contains the *coxIII* coding sequence adjacent to the *cob* coding region by a combination of block J and block E (Fig. 5E). *Cob* has two in frame start codons (Fig. 5F). Using rapid amplification of cDNA ends (RACE), we showed that the 5'-end of the *cob* mRNA is 6 nt upstream of the second start codon and 10 nt downstream of RNA19 (Fig. 5 F,G). Being only four nucleotides apart, the 3'-end of RNA19 is possibly generated during the formation of the 5'-end of *cob*. The RNA19 sequence is conserved in other apicomplexan mitochondrial genomes (Fig. 5E) but is not part of open reading frames in *Plasmodium*. Homologies and a structural fit to the ribosomal RNA of the large subunit of *E. coli* ribosomes have been noted for *P. falciparum* RNA19 (Fig. 5E, Feagin et al. 2012). This sequence similarity to rRNA is maintained in *T. gondii* (Fig. 5E), which suggests that despite overlapping with two coding sequences at the J-E block border, RNA19 is functional. It is remarkable that this sequence serves dual purposes for coding a protein and an rRNA. There is only low sequence conservation of the c*oxIII* and the *cob* sequence used by RNA19 (Fig. 5F), which might have allowed the evolutionary acquisition of an additional use as an rRNA fragment. In addition to RNA19, a second RNA encompassing coding sequences is RNA3, which is situated at the S-R block border (Fig. 5B). Block S is part of the *coxI* coding region - its terminal 22 nucleotides contribute to RNA3. The similarity to rRNA is, however, restricted to the noncoding block R (Fig. S7). The Block S sequence in RNA3 is therefore a 5'-extension of this rRNA, and it remains to be determined whether this is of functional consequence for the mitochondrial ribosome.

In sum, these results suggest that at block borders, *T. gondii* makes dual use of several protein-coding sequence blocks. Depending on the respective block combination, they are either part of the protein coding sequence or can code for an rRNA fragment.

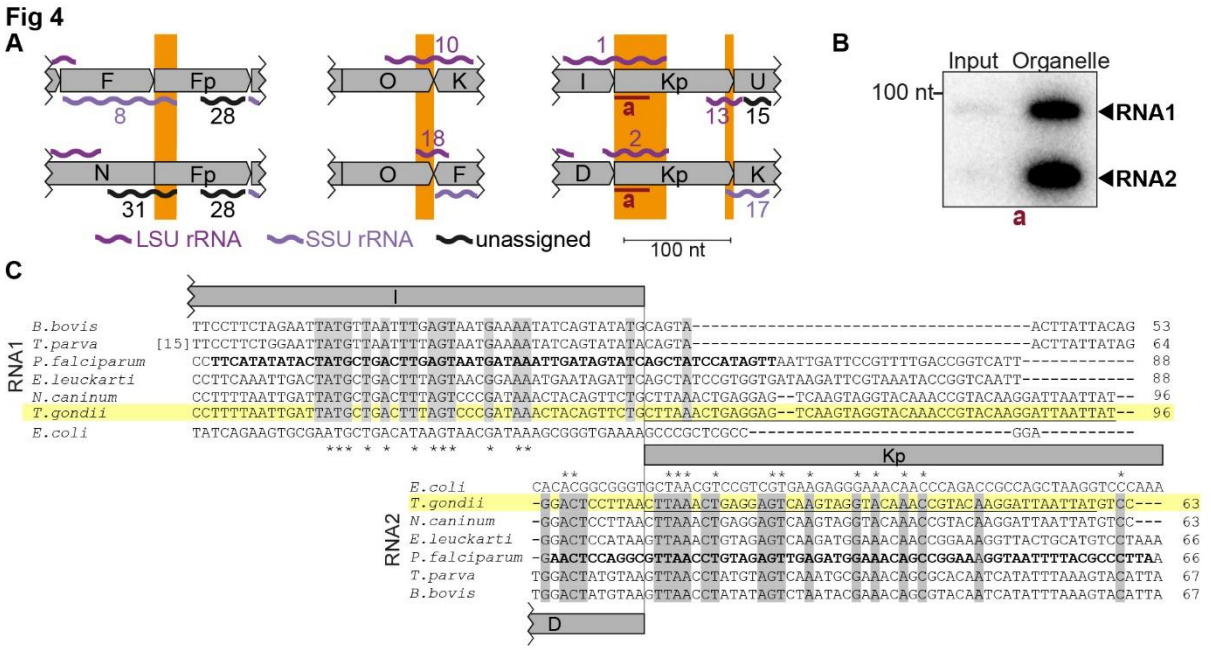

Figure 4: Noncoding RNAs at block borders.

A) Schematic representation of three block borders that are used by different transcripts (sequences used for two transcripts are marked by orange rectangles). Transcripts are shown as wavy lines and are named according to prior nomenclature (Namasivayam et al. 2021). Assignment of transcripts to the large and the small rRNA are indicated by color. The red bars with the lower-case letter "a" indicate the position of the probe used for detecting RNA1 and RNA2 by RNA gel blot hybridization. L/S= LSU/SSU = large/small subunit of the ribosome.

B) RNA gel blot hybridization of RNA1 and RNA2 using RNA extracts from input and organelle-enriched fractions of the organelle enrichment protocol shown in Fig. 1. Equal RNA amounts were loaded.

C) Alignment of RNA1 (top) and RNA2 (bottom) sequences from different apicomplexan species. Sequence excerpts from *E. coli* 23S rRNA (1240-1301 for RNA1, 944-1010 for RNA2) corresponding to sections of the alignment were added manually based on previous analyses (Feagin et al. 2012). Nucleotide positions conserved in all apicomplexans are shaded in gray. The residues in the *P. falciparum* sequence suggested to replace sequence elements in the *E. coli* 23S rRNA structure (Feagin et al. 2012) are marked in bold. Residues conserved between apicomplexans and *E. coli* are marked with asterisks. Note that RNA1 and RNA2 are registered according to their shared sequence in genomic block Kp.

Fig 5

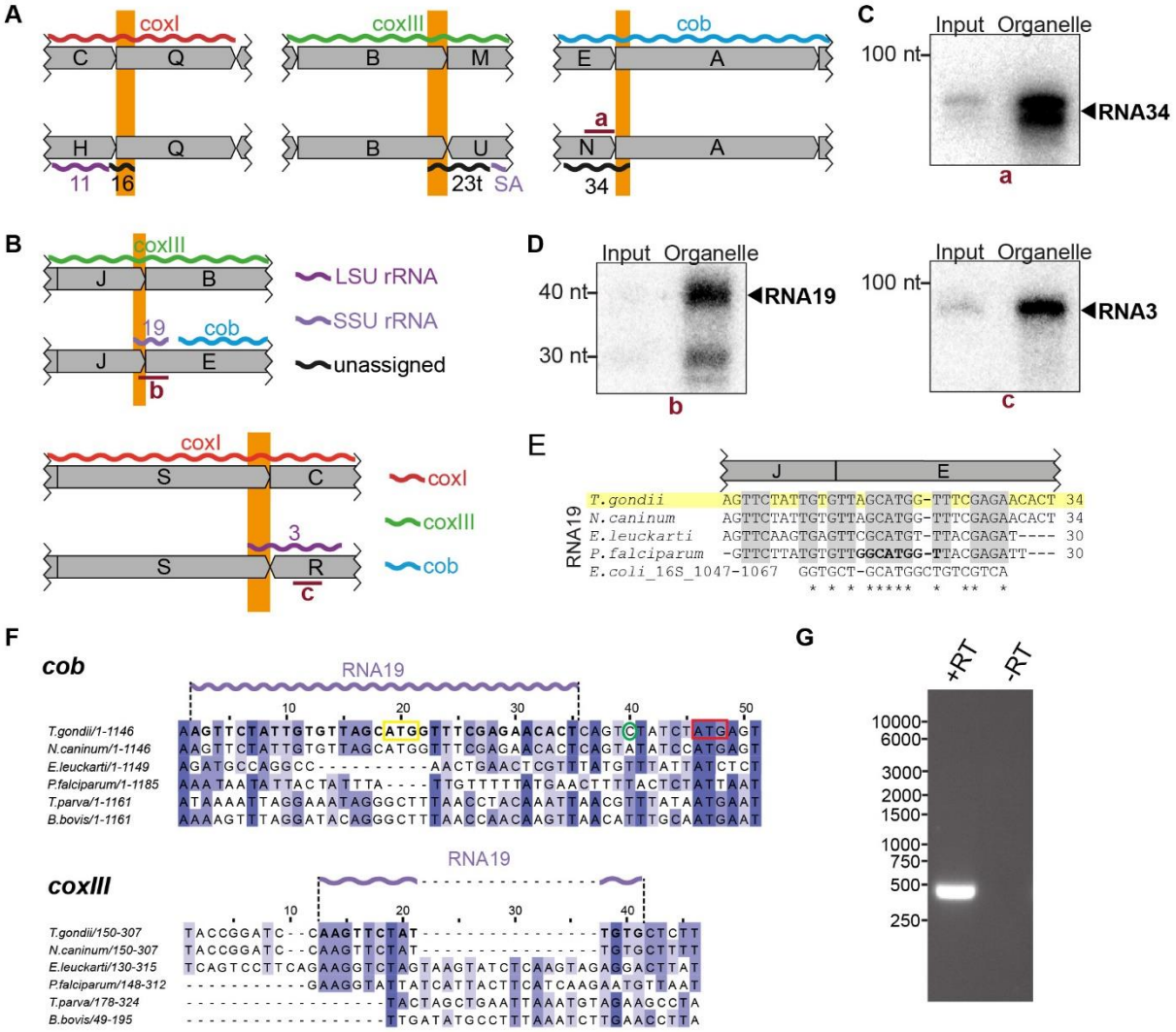

Figure 5: Small RNAs that consist of coding and noncoding sequences at block borders.

A) Schematic representation of the genomic position of three RNAs that are partially antisense to coding sequences.

B) Schematic representation of the genomic position of two RNAs that overlap with coding sequences.

C) RNA gel blot hybridization of RNA34. Equal amounts of RNA extracts from input and organelle-enriched fractions based on the organelle enrichment protocol shown in Fig. 1 were analyzed.

D) RNA gel blot hybridization of RNA19 and RNA3 – details see (B)

E) Alignment of *T. gondii* RNA19 with homologous sequences from other apicomplexan species and from *E. coli*.

F) Alignment of a section of the *coxIII* and *cob* sequences from different apicomplexan species. Alignments were prepared using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Nucleotides of *T. gondii* RNA19 are marked in bold in both alignments and as a lilac wiggle line above the sequences. The sequences are interrupted by indels in *coxIII* and *cob* of other apicomplexans that have stand-alone RNA19 sequences (e.g. *P. falciparum*). Within the *cob* sequence, the RNA19 sequence covers an annotated start codon (Alday et al. 2017; yellow box, Namasivayam et al. 2021). However, an alternative start codon downstream has been suggested as well (red box, McFadden et al.

2000) and is in fact where sequence conservation starts. The green circle in indicates the *cob* 5' transcript terminus as determined by RACE (see G).
G) Agarose gel analysis of 5' RACE products performed to identify the *cob* 5' transcript end. +RT= RACE-PCR with *T. gondii* cDNA as template, -RT= control experiment without reverse transcription

Mitochondrial small RNAs are part of a large-molecular weight complex

The apicomplexan mitochondrial ribosome with its peculiar organization has not been characterized in any detail. There is strong evidence from genetic and pharmacological studies that translation occurs in the mitochondria of apicomplexans (Alday et al. 2017; Vaidya et al. 1993; McFadden et al. 2000; Lane et al. 2018). Additionally, the presence of a large Megadalton complex containing nuclear-encoded ribosomal proteins targeted to mitochondria has been shown by blue native gel electrophoresis (Lacombe et al. 2019). However, a comprehensive catalog of mitoribosomal constituents is missing, and the extent to which rRNA fragments are part of mitoribosomes remains unanswered. We selected three short RNAs to test for their presence in high-molecular weight complexes by sucrose gradient analysis of enriched mitochondria (Lacombe et al. 2019; Shikha et al. 2022). We picked sRNAs that were well detected in RNA gel blot experiments (Fig. 4, 5, S5). LSUF/G was selected since it has strong sequence similarities to *E. coli* 23S rRNA (Feagin et al. 2012), is found in all apicomplexan mitochondrial genomes and is one of the longest rRNA fragments. RNA6 was chosen as a much shorter sRNA (56 nt). RNA3 was investigated since it is one of the sRNAs at a block boundary, including the *cox1* sequence and sequence from the noncoding block R. We separated total cellular RNA by sucrose density gradient centrifugation and detected short RNAs by RNA gel blot hybridization. As a positive control, we probed for an rRNA species of the large subunit of the cytosolic ribosome, the 5.8S rRNA. 5.8S migrated deep into the gradient, indicative of polysome association (Fig. 6). The three short RNAs were detected in fractions 5 to 12, with weaker signals present in fractions 13 and 14 (Fig. 6). This deep penetration into the gradient suggests polysome association. All three sRNAs showed similar signal distributions, peaking in fractions 7 and 8. This similarity is in line with all three of them being assigned to the 23S rRNA of *E. coli* and thus their expected identical distribution pattern

with the large subunit of the ribosome. The signal ratio between higher fractions 7 and 8, presumably monosomes, and deeper fractions ≥ 9, presumably polysomes, is different for mitochondrial rRNAs and the cytosolic 5.8S rRNA. Whether this finding has technical reasons or is explained by less dense ribosome packing on mitochondrial mRNAs versus cytosolic mRNAs cannot be determined at present. In any case, our data suggest that RNAs 3,6 and LSUF/G are part of polysomes and are thus *bona fide* ribosomal RNAs.

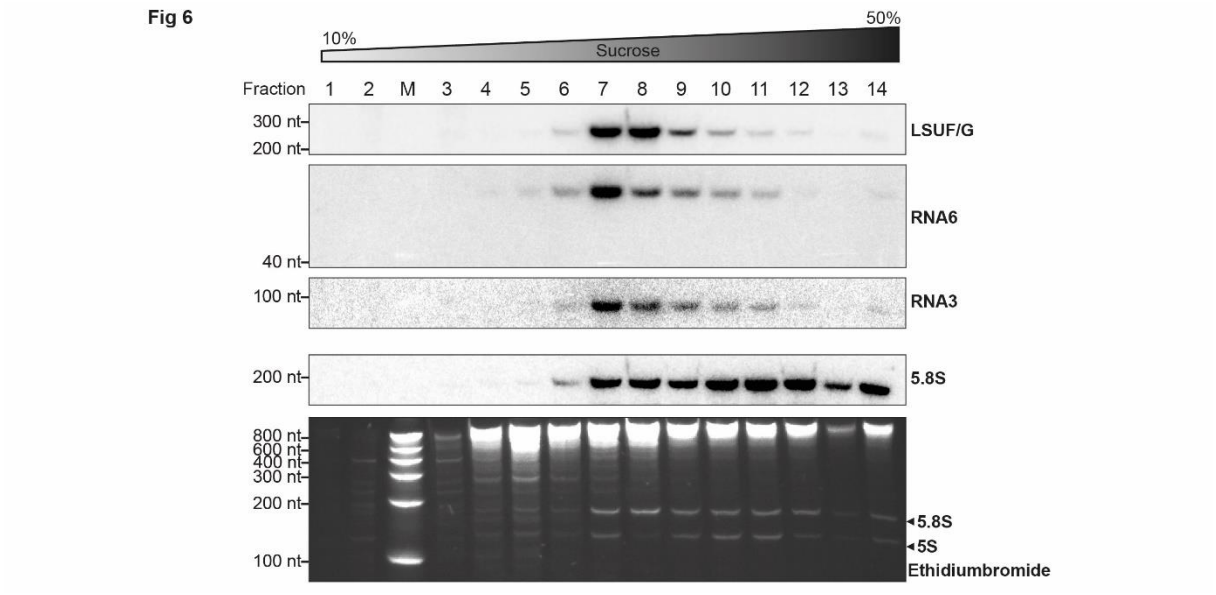

Figure 6: Sucrose density centrifugation analysis demonstrates the association of mitochondrial small RNAs with a large molecular weight complex.

Total RNA from *T. gondii* cell extracts fractionated by sucrose density gradient centrifugation was analyzed by RNA gel blot hybridization. M = size marker lane. Top: schematic representation of the sucrose density gradient across the fractions used. Middle panels: autoradiographs after probing an RNA gel blot using radiolabeled DNA probes against the short RNAs indicated on the right. Probes were used consecutively on the same blot after prior stripping of the previous probe. Bottom: ethidium bromide staining of the gel prior to blotting. Cytoplasmic 5.8S and 5S rRNAs are indicated on the right.

Discussion

Constraints in the recombination-active mitochondrial DNA of Toxoplasma gondii

The complexity of the *T. gondii* mitochondrial genome is puzzling, with an extensive variety of sequence block combinations and repetitions. The individual ONT reads are not repeated in the two published datasets (Namasivayam et al. 2021; Berná et al. 2021a) or in the data presented here, suggesting that continuous recombination shuffles the blocks. At first sight,

this might be considered a sign of evolutionary tinkering, similar to the mitochondrial idiosyncrasies found in other taxa that have been suggested to be non-adaptive, such as the fragmented mitochondrial genomes of diplonemids (Burger and Valach 2018) or the repetitive and heterogeneous genomes of plant mitochondria (Kozik et al. 2019). Upon closer scrutiny, however, the reshuffling appears limited to specific block borders and is not random. In fact, the improved sequencing depth presented in this study revealed strong constraints on the allowed block combinations in *T. gondii* mtDNA. Thus, a limited number of allowed recombination sites generates the variety of actual DNA fragments sequenced. The question remains whether the constraints outlined in the comprehensive model of mitochondrial genome architecture presented here indicate functional features.

One intriguing finding is the obligate linkage of coding sequences in a head-to-head and head-to-tail fashion. This linkage could potentially aid concerted gene regulation by utilizing shared promoter sequence elements or shared sequences for RNA stability and translation. There is evidence for coregulation of mitochondrial reading frames in other eukaryotes, including balancing the expression of *coxI* and *coxIII* and of *coxI* and *cob* in yeast (Chagas et al. 2022; Soto et al. 2012). Coregulation is not only energy-efficient but also critical since *coxI* has a metal prosthetic group that can generate reactive oxygen species in the case of protein overproduction. Linking mRNA and protein levels for *cob*, *coxI*, and *coxIII* could be facilitated by joint usage of regulatory regions.

The purpose of genome blocks in light of small RNA production

Our small RNA sequencing results revealed another potential advantage of the block organization of the *T. gondii* mitochondrial genome: We discovered 15 small RNAs that span block borders. There are different explanations for this peculiar localization of small RNAs. One possibility is that these RNAs are involved in the DNA recombination and replication process that occurs at block borders. RNA molecules have been found to play important roles in DNA repair at double-stranded breaks of nuclear DNA, both as long and small RNAs (Wei

et al. 2012; Ohle et al. 2016). However, it is currently unclear whether RNA:DNA hybrids play a role in mitochondrial DNA recombination and repair (Allkanjari and Baldock 2021).

Another possible explanation for the presence of small RNAs at block borders is that *T. gondii* simply utilizes all available genomic space for RNA production, including block borders. *T. gondii* appears to be as economical with its mitochondrial sequence space as other apicomplexans, and it does have hardly any unused noncoding sequences in its genome. There is evidence from recent polyA-sequencing efforts that long transcripts spanning several short RNAs are also produced in *T. gondii* (Lee et al. 2021). This is reminiscent of the situation in *P. falciparum*, where both strands of the entire mitochondrial genome are transcribed into polycistronic precursor RNAs. Therefore, block borders in *T. gondii* are likely transcribed into RNA as well. The functionality of these block border RNAs is unclear, but the unique sequence combinations generated at the block borders give rise to sRNAs that are not found in other apicomplexans with a simpler genome organization.

The discovery of sRNAs sharing sequences with mRNAs or being antisense to mRNAs in *T. gondii* mitochondria is an unexpected finding of our study that was facilitated by our improved understanding of mitochondrial genome organization. Antisense RNA is typically removed in mitochondria from flies and humans (Pajak et al. 2019). Similar antisense degradation mechanisms must be assumed in apicomplexans, as both strands of their genomes are transcribed in full (Ji et al. 1996; Rehkopf et al. 2000). The fact that three small antisense RNAs survive antisense RNA surveillance in *T. gondii* mitochondria is intriguing, and it remains to be established whether RNA:RNA interactions of these sRNAs with mRNAs occur *in vivo* and could be of regulatory impact.

Another potential regulatory link between mRNAs and sRNAs is represented by small RNAs that utilize coding regions. *RNA19* is of particular interest because it contains a piece of *coxIII* at the 5'-end and terminates close to the *cob* start codon. while also showing homology to rRNA. A few cases of mRNA sequences overlapping in antisense orientation with rRNA have been described in mammals and yeast (Kermekchiev and Ivanova 2001; Coelho et al. 2002).

Sequences homologous to rRNAs have been found in many coding regions in sense and antisense orientation (Mauro and Edelman 1997). It has been suggested that this could link ribosome production to other cellular processes by reciprocal inhibition of mRNA and rRNA expression (Coelho et al. 2002). It is possible that *RNA19* and ribosome production could be balanced with *cob* protein production by tuning the processing or RNA degradation of *cob* mRNA. A quantitative analysis of RNA19 and *cob* mRNA accumulation under different conditions could help to clarify whether there is such an inverse correlation. Overall, the discovery of block-border sRNAs highlights the complex regulatory mechanisms at play in *T. gondii* mitochondria and will be a starting point to understand the processing of sRNAs and their function in general.

sRNAs are incorporated into polysome-size structures

A major unsolved problem in apicomplexan mitochondrial gene expression is the nature of the ribosome. Recently, compelling evidence has surfaced suggesting the existence of ribosome-sized particles, including ribosomal proteins. Using blue-native gel electrophoresis, it was demonstrated that tagged ribosomal proteins TgmS35, TgbL12m, and TguL3m migrate as constituents of a macromolecular complex within the size range of a ribosome (Lacombe et al. 2019). Additionally, TgmS35 coprecipitated with mitochondria-encoded rRNA. Knockdown of TgmS35 resulted in specific loss of activity of one respiratory complex, which is partially encoded in the mitochondrial genome (Lacombe et al. 2019). Nevertheless, it remains to be determined whether these large complexes are actively involved in translation.

In our study, we take a step further and provide evidence that sRNAs with homology to rRNA are part of even larger complexes, most likely polysomes. This finding strongly supports the notion of active translation by mitochondrial ribosomes in tachyzoites, thus dispelling any doubts about whether rRNA fragments are part of the apicomplexan mitochondrial ribosome. Notably, even the RNA6 molecule, which is only 56 nucleotides long, migrates with polysomes, along with the much longer LSUF/G RNA. Furthermore, we demonstrate the incorporation of

RNA3, located at a block border, into polysomes, suggesting that the block border RNAs identified in this study might represent functional RNA species in general. Nonetheless, the exact role and mechanism by which sRNA fragments integrate into the ribosome for translation remain unanswered.

The structure of mitochondrial ribosomes exhibits significant variation across organisms (Tobiasson et al. 2022; Waltz et al. 2021b, 2020; Ramrath et al. 2018). In particular, mitoribosomes present an extreme example of ribosome diversification. The comprehensive collection of small RNA fragments identified in our study, including those at block borders, will serve as a valuable resource for future investigations aimed at elucidating the structure and function of the *T. gondii* mitochondrial ribosome.

Conclusions

Apicomplexan mitochondrial genomes are vital for organism survival, expressing key respiratory chain and gene expression components, particularly ribosomes. The latter were proposed to be assembled from highly fragmented ribosomal RNAs, but whether rRNA fragments are expressed and used in ribosomes is unclear. We developed a protocol to enrich *T. gondii* mitochondria and used Nanopore sequencing to comprehensively map the genome with its repeated sequence blocks. Small RNA sequencing identified fragmented ribosomal RNAs, including some RNAs spanning block boundaries, thus fusing protein-coding and rRNA sequences. Sucrose density gradient analysis showed that such rRNA fragments are in polysome-size complexes. *T. gondii's* block-based genome organization allows mitochondrial

sequences to function as both mRNAs and rRNAs, potentially linking rRNA and mRNA expression regulation.

Methods

Cultivation of host cells and *T. gondii* parasites

T. gondii parasites were cultured according to standard procedures (Jacot et al. 2020) unless otherwise indicated. Briefly, human foreskin fibroblasts (HFF-1, ATCC SCRC-1041™) were grown as host cells in Dulbecco's modified Eagle's medium (DMEM) (Capricorn Scientific) supplemented with 10% fetal bovine serum (FBS) (Capricorn Scientific) and 100 µg/ml penicillin–streptomycin (Capricorn Scientific). *T. gondii* tachyzoites were maintained in flasks containing confluent HFF-1 cells in DMEM supplemented with 1% FBS and 100 µg/ml penicillin–streptomycin. All experiments were carried out with parasites of the strain RH pSAG1-βGal/pmt-GFP, kindly provided by Frank Seeber. This strain consists of RHβ1 (PMID: 8635747), which in addition expresses eGFP N-terminally fused to the *T. gondii* mitochondrial targeting sequence S9(33-159) (DeRocher et al. 2000).

T. gondii organelle enrichment

A previously established protocol to enrich *T. gondii* organelles was modified here slightly (Esseiva et al. 2004). Freshly lysed *T. gondii* cultures were filtered through a 3 µm pore size polycarbonate filter to remove host cell debris and harvested by centrifugation at 1500 x *g* for 10 minutes. The parasite pellet was briefly washed in ice-cold Dulbecco's phosphate-buffered saline (PBS) (Capricorn Scientific) and again centrifuged for 10 minutes at 1500 x *g*. Afterwards, the parasite pellet was resuspended in 20 mM Tris-HCl, pH 7,4, 600 mM mannitol, 5 mM $MgCl_2$, and 2 mM EDTA containing 0,05% digitonin (Invitrogen) and incubated for 5 minutes on ice. The extract was clarified by centrifugation at 8000 x *g* for 5 minutes at 4°C. The supernatant containing soluble cytosolic components was discarded, and the corresponding pellet was resuspended in 20 mM Tris-HCl, pH 7.4, 600 mM mannitol, 5 mM $MgCl_2$, and 1 mM EDTA containing benzonase (345 U/ml) (Novagen), incubated for 20 minutes

at room temperature and subsequently centrifuged at 8000 x *g* for 5 minutes at 4°C. The pellet was retained and washed in 20 mM Tris-HCl, pH 7.4, 600 mM mannitol, and 17 mM EDTA to inactivate Benzonase remnants. Following a spin at 8000 x *g* for 5 minutes at 4°C, the pellet consisted of a crude enrichment of organelles and membranes. To study the *T. gondii* mitochondrial DNA content, the organelle pellet was directly used for DNA isolation. For organellar RNA analysis, the organelle pellet was frozen at -80°C for at least 24 hours. Afterwards, it was thawed on ice and resuspended in a high detergent lysis buffer (20 mM HEPES-KOH pH 7,5, 100 mM KCl, 30 mM $MgCl_2$, 1% Igepal-CA630, 1,5% Triton X-100, 0,5% sodium deoxycholate, 1 mM DTT, EDTA-free protease inhibitor). The lysate was incubated during continuous end-over-end rotation at 4°C for 40 minutes and afterwards clarified by centrifugation at 16000 x *g* for 10 minutes at 4°C. The pellet, composed of cell debris and membranes, was discarded, and the supernatant, which contained the organelle lysate, was resuspended in TRIzol® Reagent (Invitrogen) for RNA isolation.

SDS–PAGE and immunoblotting

To detect mitochondrial GFP in RH pSAG1-βGal/pmt-GFP parasites, samples taken during organelle enrichment were directly diluted in Laemmli buffer (0.2 M Tris–HCl, 8% (w/v) SDS, 40% glycerol, 20% (v/v) β-mercaptoethanol, 0.005% bromophenol blue), incubated at 95°C for 5 minutes and subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). SDS–PAGE and immunoblotting were carried out as described previously (Kupsch et al. 2012). For immunodetection, monoclonal primary mouse anti-GFP antibody (G1546, Sigma–Aldrich) and polyclonal rabbit anti-heat shock protein 70 antibody (AS05 083A, Agrisera) followed by secondary horseradish peroxidase-labeled goat anti-mouse antibody (ab205719, Abcam) and goat anti-rabbit antibody (ab205718, Abcam) were used.

DNA isolation and library construction

DNA was isolated from organelle-enriched fractions using the GeneMATRIX Tissue DNA Purification Kit (Roboklon) according to the manufacturer's protocol for cultured cells. The

Oxford Nanopore Technology sequencing library was prepared with 1 µg of organelle-enriched DNA using the SQK-LSK108 ligation sequencing kit (Oxford Nanopore, ONT) according to the manufacturer's instructions. Sequencing was performed on an R9.4.1 Spot ON Flow Cell; live-basecalling was done using the ONT Guppy software package.

RNA isolation and RNA Seq library construction

RNA was isolated from total organelle enriched and sucrose gradient fraction samples using TRIzol® Reagent (Invitrogen), followed by the Monarch™ Total RNA Miniprep Kit (New England BioLabs) applying the manufacturer's protocol for TRIzol® extracted samples. Two hundred nanograms of RNA was used to generate sequencing libraries using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® Kit (New England BioLabs) according to the manufacturer's instructions. For identification of PCR duplicates in the library preparation process, the 5´-linker was modified with 8 N-nucleotides (as unique molecular identifiers, UMIs). Small RNA library clean-up was carried out with AMPure XP Beads (Beckmann Coulter). Sequencing was performed by Genewiz on an Illumina NovaSeq platform with a read length of 150 bp for each pair.

Agarose-formaldehyde gel electrophoresis and sRNA gel blot

RNA was separated on a 1% agarose gel containing 1.2% formaldehyde, and cytosolic rRNAs were detected with ethidium bromide staining. Small RNA gel blotting and detection of transcripts with radiolabeled DNA oligo probes was carried out as described previously (Loizeau et al. 2014). Sequences of the oligo probes used in this study are listed in Tab. S6. Blot hybridization of sucrose density gradient fractions was carried out consecutively in the

order LSUF_G/RNA6/RNA3/5.8S, with removal of probes by heating the blot to 60°C for 1 hour in 0.5% SDS.

5' RACE

T. gondii total RNA was ligated to a small RNA oligo (Rumsh, see Table S6) using T4 RNA Ligase 1 (NEB) according to the manufacturer's instructions. Afterwards, RNA was reverse transcribed using ProtoScript II Reverse Transcriptase (NEB) and random primers. PCR amplification of cDNA was performed with a gene-specific *cob* primer and a primer annealing to the ligated oligo (Rumsh1, see Table S6) using Taq DNA Polymerase (Roboklon). The purified PCR product was sequenced by Sanger Sequencing (LGC Genomics GmbH).

Sucrose density gradient centrifugation

Ribosome fractionation by sucrose density gradient centrifugation was performed as previously described (Waltz et al. 2021a). Briefly, *T. gondii* parasites were harvested and immediately frozen at -80°C. On the day of the experiment, parasites were thawed on ice and lysed in 20 mM HEPES-KOH pH 7,5, 100 mM KCl, 30 mM $MgCl_2$, 1% Igepal-CA630, 1,5% Triton X-100, 0,5% sodium deoxycholate, 1 mM DTT, and EDTA-free protease inhibitor for 40 minutes at 4°C under continuous end-over-end rotation. The lysate was cleared by centrifugation at 16000xg for 10 minutes at 4°C, and the supernatant was layered onto a continuous 10%-50% sucrose gradient. Sucrose solutions (10% and 50%) were previously prepared freshly in 20 mM HEPES-KOH pH 7.5, 100 mM KCl, 30 mM $MgCl_2$1 mM DTT, and EDTA-free protease inhibitor, and a continuous gradient was formed using BioComp Gradient Master 108. The gradient was centrifuged at 200.000xg for 2 hours at 4°C in a S52-ST rotor (Thermo Scientific) and subsequently fractionated into 14 fractions using the BioComp Piston

Gradient Fractionator device. Fractions were resuspended directly in TRIzol® reagent (Invitrogen) and subjected to RNA isolation. Paste your materials and methods section here.

Bioinformatic analyses

All bioinformatic analyses except for the quantification of sequence block combinations were performed on the Galaxy web platform (v.22.01; https://usegalaxy.eu, Afgan et al. 2018). All reference sequences used were retrieved from GenBank (Sayers et al. 2019). Sequence alignments were prepared using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with default settings. *E. coli* sequences were added manually according to sequence positions established previously (Feagin et al. 2012).

Masking NUMTs in the T. gondii nuclear genome

To identify NUMTs in the *T. gondii* nuclear genome, mitochondrial sequence blocks (MN077088.1-MN077111.1) were aligned against the RH-88 nuclear reference genome (GCA_019455545.1, August 2021) using NCBI-BLASTN (v. 2.13.0) (Altschul et al. 1990) (default parameters, except max_target_seqs = 5000). Intervals of the nuclear genome aligning to mitochondrial sequences obtained in the BLASTN tabular output were converted into GFF3 format. NUMTs in the nuclear genome were masked with bedtools MaskFastaBed (Quinlan and Hall 2010).

DNA sequencing data analysis

DNA Oxford Nanopore sequencing data were trimmed using porechop (Version 0.2.4, https://github.com/rrwick/Porechop). As a first step, trimmed reads were mapped against the human reference genome GRCh38.p14 (GCA_000001405.29, February 2022) using Minimap2 (Version 2.24) (Li 2018) (PacBio/Oxford Nanopore read to reference mapping, default settings). Mapping results were filtered using SAMtools (Version 1.15.1) (Li et al. 2009). Unmapped reads were subsequently simultaneously mapped against the *T. gondii* RH-88 nuclear reference genome masked for NUMTS, the apicoplast genome (CM033583.1) and

mitochondrial sequence blocks (MN077088.1- MN077111.1) using Minimap2. A fraction of the reads that fit best to the mitochondrial genome still retain shorter nuclear sequence sections and are therefore representing NUMTS. We therefore took all reads that mapped to the mitochondrial genome in our first mapping and mapped them against the nuclear reference genome masked for NUMTS using Minimap2 (Version 2.24). The remaining unmapped reads were considered mitochondrial and visualized with Geneious Prime® v. 2019.2.3. (Biomatters Ltd.). Reads were annotated providing the mitochondrial sequence blocks (MN077088.1 - MN077111.1) and ORFs (MN077082.1 - MN077084.1) as annotation database using the Geneious Prime® “Annotate from Database” tool with similarity set to ≥75%.

Sequence comparisons of ONT reads found here with published ONT reads for the *T. gondii* mitochondrial genome were performed using the NUCmer tool in the MUMmer package of Galaxy with the following settings: minimum match length 20 and minimum alignment length 290 (shortest read in Namasivayam et al. 2021). The Show-Coords tool was used to output alignment information. No matching reads were found.

Analysis of mitochondrial DNA sequence block arrangements

Annotated blocks were sorted by start location, transformed into strings of block IDs per read and exported in plain text format, which was used as input for further analysis in R. We excluded all reads with fewer than three blocks and extracted all two-block combinations as seen, including start and end blocks. Using a custom script employing the packages *tidyr*, *dplyr*, and *stringr*, we parsed the block ID annotation file and counted all block IDs across all reads. Then, all possible two-block combinations were simulated, and their occurrences in our read dataset were counted. The R code and input block ID order file are provided in the

Supplemental Material. The results were manually curated to exclude false-positive combinations detected due to annotation gaps caused by the error-prone nature of ONT reads.

RNA sequencing data analysis

Adapter and quality trimming of RNA sequencing data was performed using Cutadapt (Martin 2011). Umi-tools was used to extract UMIs (Smith et al. 2017). Reads were aligned to the *T. gondii* RH-88 nuclear reference genome masked for NUMTS and the apicoplast genome (CM033583.1) using Bowtie2 (Langmead et al. 2009) in default settings. Mapping results were filtered using SAMtools (Version 1.15.1) (Li et al. 2009). Unmapped reads were kept and mapped with Bowtie2 against the *T. gondii* mitochondrial genome using combinations of the sequence blocks (GenBank accession MN077088.1 - MN077111.1, OR086910 - OR086916) as determined in our ONT DNA sequencing data. The results were again processed using SAMtools. Transcripts were determined as follows: first, most transcripts had short poly-A tails - up to 12 adenines at their 3´-ends - which is similar to the situation in *P. falciparum* (Rehkopf et al. 2000). For the identification of the transcript 3'-ends, all polyA-tails longer than three nucleotides were removed from reads using fastq (Chen et al. 2018). The 3' and 5' ends of reads were mapped using Bedtools Genome Coverage 2.30.0 (Quinlan and Hall 2010). The 5'-ends at the beginning of coverage plateaus over the background were considered starting points of transcripts. For each 5' end, the mapped 3'-end most far away but still represented 10% of the reads of the 5'-terminus and connected to the starting point by uninterrupted coverage over background was considered the corresponding 3'-terminus. Alignments and coverage graphs were visualized in Integrated Genome Viewer (Robinson et al. 2011). All RNA sequencing data are available at SRA under accession number PRJNA978626.

Homology search for novel transcript annotation

All newly identified transcripts were subjected to an NCBI-BLASTN homology search (v. 2.13.0; Altschul et al. 1990; default parameters, except word size=7) against the *Plasmodium falciparum* mitochondrial genome (M76611.1) and the *Eimeria leuckarti* mitochondrial genome

(MW354691.1). Transcripts not matching the *P. falciparum* or *E. leuckarti* mitochondrial genome were further inspected for homology to the *E. coli* rRNA operon (J01695.2) using BLASTN.

Data access

The RNA and DNA sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA978626. The script for block analysis and input tables can be found at Mendeley Data, doi:10.17632/vz76z7ps2y.1. New block annotations can be found in GenBank, OR086910 - OR086916.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We are grateful to Giel van Dooren, Frank Seeber, Kai Matuschewski, and Hannes Ruwe for critical discussion of our data and manuscript.

Authors' contributions

S. W: and C.S. designed the research; S.T., A.H. and P.L. performed the research; S.T., A.H. and S.W. analyzed the data; C.S. wrote the paper with contributions from all other authors.

Funding

This work was supported by the German research foundation (DFG) by grant IRTG2290, project B01 to C.S.

References

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, et al. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**: W537–W544.

Alday PH, Bruzual I, Nilsen A, Pou S, Winter R, Ben Mamoun C, Riscoe MK, Doggett JS. 2017. Genetic Evidence for Cytochrome b Qi Site Inhibition by 4(1H)-Quinolone-3-Diarylethers and Antimycin in Toxoplasma gondii. *Antimicrob Agents Chemother* **61**. http://dx.doi.org/10.1128/AAC.01866-16.

Allkanjari K, Baldock RA. 2021. Beyond base excision repair: an evolving picture of mitochondrial DNA repair. *Biosci Rep* **41**. http://dx.doi.org/10.1042/BSR20211320.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Berná L, Marquez P, Cabrera A, Greif G, Francia ME, Robello C. 2021a. Reevaluation of the Toxoplasma gondii and Neospora caninum genomes reveals misassembly, karyotype differences, and chromosomal rearrangements. *Genome Res* **31**: 823–833.

Berná L, Rego N, Francia ME. 2021b. The Elusive Mitochondrial Genomes of Apicomplexa: Where Are We Now? *Front Microbiol* **12**: 751775.

Burger G, Valach M. 2018. Perfection of eccentricity: Mitochondrial genomes of diplonemids. *IUBMB Life* **70**: 1197–1206.

Bushell E, Gomes AR, Sanderson T, Anar B, Girling G, Herd C, Metcalf T, Modrzynska K, Schwach F, Martin RE, et al. 2017. Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell* **170**: 260-272.e8.

Chagas JAC, Kfouri Martins Soares MA, Ribeiro Franco LV, Barros MH. 2022. Overexpression of MRX9 impairs processing of RNAs encoding mitochondrial oxidative phosphorylation factors COB and COX1 in yeast. *J Biol Chem* **298**: 102214.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Coelho PSR, Bryan AC, Kumar A, Shadel GS, Snyder M. 2002. A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA. *Genes Dev* **16**: 2755–2760.

DeRocher A, Hagen CB, Froehlich JE, Feagin JE, Parsons M. 2000. Analysis of targeting sequences demonstrates that trafficking to the Toxoplasma gondii plastid branches off the secretory system. *J Cell Sci* **113 (Pt 22)**: 3969–3977.

Esseiva AC, Naguleswaran A, Hemphill A, Schneider A. 2004. Mitochondrial tRNA import in Toxoplasma gondii. *J Biol Chem* **279**: 42363–42368.

Feagin JE, Harrell MI, Lee JC, Coe KJ, Sands BH, Cannone JJ, Tami G, Schnare MN, Gutell RR. 2012. The fragmented mitochondrial ribosomal RNAs of Plasmodium falciparum. *PLoS One* **7**: e38320.

Gjerde B. 2013. Characterisation of full-length mitochondrial copies and partial nuclear copies (numts) of the cytochrome b and cytochrome c oxidase subunit I genes of Toxoplasma gondii, Neospora caninum, Hammondia heydorni and Hammondia triffittae (Apicomplexa: Sarcocystidae). *Parasitol Res* **112**: 1493–1511.

Hikosaka K, Kita K, Tanabe K. 2013. Diversity of mitochondrial genome structure in the phylum Apicomplexa. *Mol Biochem Parasitol* **188**: 26–33.

Hillebrand A, Matz JM, Almendinger M, Muller K, Matuschewski K, Schmitz-Linneweber C. 2018. Identification of clustered organellar short (cos) RNAs and of a conserved family of organellar RNA-binding proteins, the heptatricopeptide repeat proteins, in the malaria parasite. *Nucleic Acids Res*. https://www.ncbi.nlm.nih.gov/pubmed/30102371.

Hollin T, Abel S, Falla A, Pasaje CFA, Bhatia A, Hur M, Kirkwood JS, Saraf A, Prudhomme J, De Souza A, et al. 2022. Functional genomics of RAP proteins and their role in mitoribosome regulation in Plasmodium falciparum. *Nat Commun* **13**: 1275.

Jacot D, Lourido S, Meissner M, Sheiner L, Soldati-Favre D, Striepen B. 2020. Chapter 20 - Genetic manipulation of Toxoplasma gondii. In *Toxoplasma gondii* (ed. K.K. Louis M. Weiss), pp. 897–940.

Ji YE, Mericle BL, Rehkopf DH, Anderson JD, Feagin JE. 1996. The Plasmodium falciparum 6 kb element is polycistronically transcribed. *Mol Biochem Parasitol* **81**: 211–223.

Ke H, Dass S, Morrisey JM, Mather MW, Vaidya AB. 2018. The mitochondrial ribosomal protein L13 is critical for the structural and functional integrity of the mitochondrion in Plasmodium falciparum. *J Biol Chem* **293**: 8128–8137.

Ke H, Morrisey JM, Ganesan SM, Mather MW, Vaidya AB. 2012. Mitochondrial RNA polymerase is an essential enzyme in erythrocytic stages of Plasmodium falciparum. *Mol Biochem Parasitol* **185**: 48–51.

Kermekchiev M, Ivanova L. 2001. Ribin, a protein encoded by a message complementary to rRNA, modulates ribosomal transcription and cell proliferation. *Mol Cell Biol* **21**: 8255–8263.

Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. 2019. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet* **15**: e1008373.

Kupsch C, Ruwe H, Gusewski S, Tillich M, Small I, Schmitz-Linneweber C. 2012. Arabidopsis chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell* **24**: 4266–4280.

Lacombe A, Maclean AE, Ovciarikova J, Tottey J, Mühleip A, Fernandes P, Sheiner L. 2019. Identification of the Toxoplasma gondii mitochondrial ribosome, and characterisation of a protein essential for mitochondrial translation. *Mol Microbiol* **112**: 1235–1252.

Lane KD, Mu J, Lu J, Windle ST, Liu A, Sun PD, Wellems TE. 2018. Selection of Plasmodium falciparum cytochrome B mutants by putative PfNDH2 inhibitors. *Proc Natl Acad Sci U S A* **115**: 6285–6290.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lau Y-L, Lee W-C, Gudimella R, Zhang G, Ching X-T, Razali R, Aziz F, Anwar A, Fong M-Y. 2016. Deciphering the Draft Genome of Toxoplasma gondii RH Strain. *PLoS One* **11**: e0157901.

Lee VV, Judd LM, Jex AR, Holt KE, Tonkin CJ, Ralph SA. 2021. Direct Nanopore Sequencing of mRNA Reveals Landscape of Transcript Isoforms in Apicomplexan Parasites. *mSystems* **6**. http://dx.doi.org/10.1128/mSystems.01081-20.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing, Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Loizeau K, Qu Y, Depp S, Fiechter V, Ruwe H, Lefebvre-Legendre L, Schmitz-Linneweber C, Goldschmidt-Clermont M. 2014. Small RNAs reveal two target sites of the RNA-maturation factor Mbb1 in the chloroplast of Chlamydomonas. *Nucleic Acids Res* **42**: 3286–3297.

Lorenzi H, Khan A, Behnke MS, Namasivayam S, Swapna LS, Hadjithomas M, Karamycheva S, Pinney D, Brunk BP, Ajioka JW, et al. 2016. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. *Nat Commun* **7**: 10147.

MacRae JI, Sheiner L, Nahid A, Tonkin C, Striepen B, McConville MJ. 2012. Mitochondrial metabolism of glucose and glutamine is required for intracellular growth of Toxoplasma gondii. *Cell Host Microbe* **12**: 682–692.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**. http://dx.doi.org/10.14806/ej.17.1.200.

Mauro VP, Edelman GM. 1997. rRNA-like sequences occur in diverse primary transcripts: implications for the control of gene expression. *Proc Natl Acad Sci U S A* **94**: 422–427.

McFadden DC, Tomavo S, Berry EA, Boothroyd JC. 2000. Characterization of cytochrome b from Toxoplasma gondii and Q(o) domain mutations as a mechanism of atovaquone-resistance. *Mol Biochem Parasitol* **108**: 1–12.

Melo EJ, Attias M, De Souza W. 2000. The single mitochondrion of tachyzoites of Toxoplasma gondii. *J Struct Biol* **130**: 27–33.

Minot S, Melo MB, Li F, Lu D, Niedelman W, Levine SS, Saeij JPJ. 2012. Admixture and recombination among Toxoplasma gondii lineages explain global genome diversity. *Proc Natl Acad Sci U S A* **109**: 13458–13463.

Molan A, Nosaka K, Hunter M, Wang W. 2019. Global status of Toxoplasma gondii infection: systematic review and prevalence snapshots. *Trop Biomed* **36**: 898–925.

Namasivayam S, Baptista RP, Xiao W, Hall EM, Doggett JS, Troell K, Kissinger JC. 2021. A novel fragmented mitochondrial genome in the protist pathogen Toxoplasma gondii and related tissue coccidia. *Genome Res* **31**: 852–865.

Nishi M, Hu K, Murray JM, Roos DS. 2008. Organellar dynamics during the cell cycle of Toxoplasma gondii. *J Cell Sci* **121**: 1559–1568.

Oborník M, Lukeš J. 2015. The Organellar Genomes of Chromera and Vitrella, the Phototrophic Relatives of Apicomplexan Parasites. *Annu Rev Microbiol* **69**: 129–144.

Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I, Fischer T. 2016. Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* **167**: 1001-1013.e7.

Ossorio PN, Sibley LD, Boothroyd JC. 1991. Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of Toxoplasma gondii. *J Mol Biol* **222**: 525–536.

Ovciarikova J, Lemgruber L, Stilger KL, Sullivan WJ, Sheiner L. 2017. Mitochondrial behaviour throughout the lytic cycle of Toxoplasma gondii. *Sci Rep* **7**: 42746.

Pajak A, Laine I, Clemente P, El-Fissi N, Schober FA, Maffezzini C, Calvo-Garrido J, Wibom R, Filograna R, Dhir A, et al. 2019. Defects of mitochondrial RNA turnover lead to the accumulation of double-stranded RNA in vivo. *PLoS Genet* **15**: e1008240.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Ramrath DJF, Niemann M, Leibundgut M, Bieri P, Prange C, Horn EK, Leitner A, Boehringer D, Schneider A, Ban N. 2018. Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science* **362**. http://dx.doi.org/10.1126/science.aau7735.

Rehkopf DH, Gillespie DE, Harrell MI, Feagin JE. 2000. Transcriptional mapping and RNA processing of the Plasmodium falciparum mitochondrial mRNAs. *Mol Biochem Parasitol* **105**: 91–103.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.

Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2019. GenBank. *Nucleic Acids Res* **47**: D94–D99.

Seeber F, Feagin JE, Parsons M, van Dooren GG. 2020. Chapter 11 - The apicoplast and mitochondrion of Toxoplasma gondii. In *Toxoplasma gondii (Third Edition)* (eds. L.M. Weiss and K. Kim), pp. 499–545, Academic Press.

Seeber F, Ferguson DJ, Gross U. 1998. Toxoplasma gondii: a paraformaldehyde-insensitive diaphorase activity acts as a specific histochemical marker for the single mitochondrion. *Exp Parasitol* **89**: 137–139.

Shikha S, Silva MF, Sheiner L. 2022. Identification and Validation of Toxoplasma gondii Mitoribosomal Large Subunit Components. *Microorganisms* **10**. http://dx.doi.org/10.3390/microorganisms10050863.

Sidik SM, Huet D, Ganesan SM, Huynh M-H, Wang T, Nasamu AS, Thiru P, Saeij JPJ, Carruthers VB, Niles JC, et al. 2016. A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell* **166**: 1423-1435.e12.

Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499.

Soto IC, Fontanesi F, Liu J, Barrientos A. 2012. Biogenesis and assembly of eukaryotic cytochrome c oxidase catalytic core. *Biochim Biophys Acta* **1817**: 883–897.

Subczynski WK, Pasenkiewicz-Gierula M, Widomska J, Mainali L, Raguz M. 2017. High Cholesterol/Low Cholesterol: Effects in Biological Membranes: A Review. *Cell Biochem Biophys* **75**: 369–385.

Suplick K, Morrisey J, Vaidya AB. 1990. Complex transcription from the extrachromosomal DNA encoding mitochondrial functions of Plasmodium yoelii. *Mol Cell Biol* **10**: 6381–6388.

Szabo EK, Finney CAM. 2017. Toxoplasma gondii: One Organism, Multiple Models. *Trends Parasitol* **33**: 113–127.

Tobiasson V, Berzina I, Amunts A. 2022. Structure of a mitochondrial ribosome with fragmented rRNA in complex with membrane-targeting elements. *Nat Commun* **13**: 6132.

Vaidya AB, Lashgari MS, Pologe LG, Morrisey J. 1993. Structural features of Plasmodium cytochrome b that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones. *Mol Biochem Parasitol* **58**: 33–42.

Vercesi AE, Rodrigues CO, Uyemura SA, Zhong L, Moreno SN. 1998. Respiration and oxidative phosphorylation in the apicomplexan parasite Toxoplasma gondii. *J Biol Chem* **273**: 31040–31047.

Waltz F, Giegé P, Hashem Y. 2021a. Purification and Cryo-electron Microscopy Analysis of Plant Mitochondrial Ribosomes. *Bio Protoc* **11**: e4111.

Waltz F, Salinas-Giegé T, Englmeier R, Meichel H, Soufari H, Kuhn L, Pfeffer S, Förster F, Engel BD, Giegé P, et al. 2021b. How to build a ribosome from RNA fragments in Chlamydomonas mitochondria. *Nat Commun* **12**: 7176.

Waltz F, Soufari H, Bochler A, Giegé P, Hashem Y. 2020. Cryo-EM structure of the RNA-rich plant mitochondrial ribosome. *Nat Plants* **6**: 377–383.

Wei W, Ba Z, Gao M, Wu Y, Ma Y, Amiard S, White CI, Rendtlew Danielsen JM, Yang Y-G, Qi Y. 2012. A role for small RNAs in DNA double-strand break repair. *Cell* **149**: 101–112.