BMC Genomics

RESEARCH

Open Access

Mitochondrial transcriptome of *Candida albicans in flagranti* — direct RNA sequencing reveals a new layer of information

Jakub Piątkowski[1*], Kacper Koźluk[1] and Paweł Golik[1,2]

Abstract

Background Organellar transcriptomes are relatively under-studied systems, with data related to full-length transcripts and posttranscriptional modifications remaining sparse. Direct RNA sequencing presents the possibility of accessing a previously unavailable layer of information pertaining to transcriptomic data, as well as circumventing the biases introduced by second-generation RNA-seq platforms. Direct long-read ONT sequencing allows for the isoform analysis of full-length transcripts and the detection of posttranscriptional modifications. However, there are still relatively few projects employing this method specifically for studying organellar transcriptomes.

Results *Candida albicans* is a promising model for investigating nucleo-mitochondrial interactions. This work comprises ONT sequencing of the *Candida albicans* mitochondrial transcriptome along with the development of a dedicated data analysis pipeline. This approach allowed for the detection of complete transcript isoforms and posttranslational RNA modifications, as well as an analysis of *C. albicans* deletion mutants in genes coding for the 5′ and 3′ mitochondrial RNA exonucleases CaPET127 and CaDSS1. It also enabled for corrections to previous studies in terms of 3′ and 5′ transcript ends. A number of intermediate splicing isoforms was also discovered, along with mature and unspliced transcripts and changes in their abundances resulting from disruption of both 5′ and 3′ exonucleolytic processing. Multiple putative posttranscriptional modification sites have also been detected.

Conclusions This preliminary work demonstrates the suitability of direct RNA sequencing for studying yeast mitochondrial transcriptomes in general and provides new insights into the workings of the *C. albicans* mitochondrial transcriptome in particular. It also provides a general roadmap for analyzing mitochondrial transcriptomic data from other organisms.

Keywords Direct RNA-seq, Long-read sequencing, Mitochondria, Transcriptome, *C. albicans*

Background

Yeasts, specifically *Candida albicans*, serve as a promising model for studying the coevolution of nuclear and mitochondrial genomes, as well as the regulation of mitochondrial gene expression [1]. Diverging evolutionarily from the commonly studied baker's yeast *Saccharomyces cerevisiae* by 480 million years, *C. albicans* provides valuable comparative insights [2], notably due to the presence of Complex I in its respiratory chain — a feature absent in *S. cerevisiae* but present in humans [3]. This makes *C.*

*Correspondence:
Jakub Piątkowski
j.piatkowski@uw.edu.pl
[1] Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, 02-106 Warsaw, Poland
[2] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland

albicans a preferable model for investigating the workings of the respiratory chain in human mitochondria.

The reduction of the mitochondrial genome, including regulatory elements within mtDNA, and the extensive simplification of mitochondrial transcriptional machinery (with monomeric RNA polymerase of phage origins responsible for transcript generation) in most eukaryotic organisms imply that a significant portion of gene expression regulation in mitochondria occurs posttranscriptionally [4, 5]. Therefore, it is crucial to develop methods that provide a reliable insight into the mitochondrial transcriptome and changes that occur following disruptions in RNA processing and posttranscriptional regulation.

The rapid development of high-throughput next-generation sequencing in the last decade has enabled the study of entire transcriptomes at both the cellular and organelle levels. However, commonly used second-generation sequencing technologies such as Illumina and — to a lesser extent — Ion Torrent, while generating a wealth of valuable data, are not without significant drawbacks. The process of transcribing RNA into cDNA can introduce numerous artifacts, related to the varied affinity of random oligonucleotides used in reverse transcription to specific molecules of native RNA, as well as errors occurring during the reaction itself (e.g. substitution of the synthesized strand between templates of similar sequences) [6]. Library preparation for sequencing can also lead to the enrichment of sequences corresponding to RNA molecules in a specific size range. The short length of reads likewise hinders the quantitative and qualitative analysis of different variants of the same transcript present in the cell or organelle. Additionally, during transcription into cDNA, all information regarding covalent RNA chain modifications is lost [7].

Oxford Nanopore Technologies (hereafter refered to as ONT) direct RNA sequencing overcomes these drawbacks, despite higher sequencing costs per base pair. It allows for the direct sequencing of individual RNA molecules in their entirety and the detection of a number of nucleotide modifications (such as m6A, m7G or Ψ) [8–10]. The lower throughput, compared to that of the Ion Torrent and Illumina platforms, does not significantly impact the study of smaller transcriptomes, such as the mitochondrial transcriptome. However, relatively few mitochondrial transcriptomes so far have been investigated in this manner [11, 12].

The goal of this study is to examine the mitochondrial transcriptome of *C. albicans* using the MinION sequencer (ONT), with particular attention given to 5' and 3' ends, splicing isoforms and posttranscriptional modifications. The study encompasses both wild-type strain and strains with disruptions in CaDSS1 (encoding a subunit of mtEXO complex) and the gene encoding the Pet127 protein, which is involved in processing the 3' and 5' ends of mitochondrial transcripts, respectively. Disruption of mtEXO results in a distinct growth phenotype and significant changes in the mitochondrial transcriptome [13], while deletion of CaPET127 preserves the full functionality of the respiratory chain, with only subtle changes at the transcriptome level [14]. As both these exonucleases have been well-studied in *S. cerevisiae* as well as in *C. albicans*, and found to play an important role in posttranscriptional processing (which itself plays a proportionally larger role in mitochondrial expression systems, relatively to nuclear ones) [15], analyzing them via direct RNA sequencing should yield further insights into the mitochondrial transcriptome as well as facilitate the optimization of methodology and assess the sensitivity of such an approach and the scientific value of data thus obtained. Additionally, to date no information regarding mitochondrial RNA modifications in *C. albicans* is available, and any new results in this area would be of scientific value.

Materials and methods

C. albicans strains and growth conditions

BWP17 [16] (arg4::hisG/arg4::hisG, his1::hisG/his1::hisG, ura3::imm434/ura3::imm434, iro1::imm434/iro1::imm434).

ΔCaDSS [13] (arg4/arg4, his1::hisG/his1::hisG, ura3::imm434 /ura3::imm434, iro1::imm434 /iro1::imm434, dss1::HIS1/dss1::SAT).

ΔCaPET127 [14] (arg4/arg4, his1::hisG/his1::hisG, ura3::imm434/ura3::imm434, iro1::imm434/iro1::imm434, pet127::HIS1/pet127::SAT).

The *C. albicans* strains listed above were grown in YPGal medium (1% yeast extract, 2% peptone and 2% galactose) containing 80 μg/ml uridine at 37°C until logarithmic growth phase of OD (optical density at 600 nm wavelength) between 1.6 and 2).

Mitochondria isolation

Mitochondria were isolated from log-phase liquid culture by differential centrifugation as described previously [17]. Purified intact mitochondria were treated with 10 μg of RNase A (Thermo Scientific) at 37 °C for 10 min to remove copurified cytoplasmic RNA.

RNA isolation

Isolation of mitochondrial RNA was performed by hot phenol method [18] followed by purification with Monarch® RNA Cleanup Kit (New England Biolabs). RNA was quantified with NanoDrop 2000 spectrophotometer and subsequently treated with DNase I (Thermo Scientific) with RiboLock RNase Inhibitor (Thermo Scientific)

for 15 min in 37 °C, after which the RNA was purified again with Monarch® RNA Cleanup Kit. RNA quality and concentration were then assessed with Agilent RNA 6000 Nano kit.

In vitro RNA synthesis

In vitro transcription was performed on PCR-amplified templates corresponding to all exonic mitochondrial sequences with HiScribe® T7 High Yield RNA Synthesis Kit (New England Biolabs). PCR reactions were performed on *C. albicans* mtDNA template with the primers listed in Supplementary Data. Transcribed RNA products were subsequently treated with DNase I, as described above, purified with Monarch® RNA Cleanup Kit (New England Biolabs), assessed through agarose gel electrophoresis, quantified with NanoDrop 2000 spectrophotometer and mixed in an equimolar manner. The primers used for template amplification are listed in Supplementary Table 1.

RNA demethylation with FTO

Prior to polyadenylation, 3 µg of RNA isolated from BWP17 mitochondria was incubated for 18 h at room temperature with 1 µg of FTO in FTO assay buffer (FTO Chemiluminescent Assay Kit, BPS Bioscience) with RiboLock RNase Inhibitor (Thermo Scientific). The final reaction volume was 20 µl.

LC–MS/MS and dot-blot detection of m6A

One µg of RNA (with and without FTO treatment) was digested with P1 nuclease (New England Biolabs) and dephosphorylated with FastAP phosphatase (Thermo Scientific). Samples were subsequently measured on Orbitrap Exploris 480 spectrometer with unmodified adenosine and m6A as standards [19, 20].

Dot-blot m6A immunodetection was performed as described previously [21], with anti-m6A antibody (BPS Bioscience) and with TBST used in place of PBS for buffer preparation.

Bisulfite RNA sequencing

Prior to polyadenylation, 3 µg of RNA isolated from BWP17 mitochondria or obtained via in vitro transcription was treated with EZ RNA Methylation Kit (Zymo Research) [22]. Input RNA was divided into four conversion reactions 750 ng each, for optimal conversion efficiency and recovery.

Mitochondrial RNA ribodepletion

Ribodepletion was performed following a previously described two-step protocol [23], with minor modifications. 2X SSC and 0.1% Tween-20 were used as hybridization buffer. 120 µl of DynaBeads were washed 2× in solution A and 2× in solution B. Half of the beads were resuspended in 2xB&W buffer and 10 µl of RNase-free water and 10 µl of 1xRiboHyb solution were added (for Round 1 of ribodepletion). The other half of the beads were resuspended in 120 µl in 1xB&W buffer (for Round 2 of ribodepletion). The beads were stored at 37°C. 20 µl of 2xRiboHyb solution were mixed with 5 µg of RNA and 1 µl of probe mix (10µM). The volume was brought to 40 µl with mRNase-free water.

The RNA/probe mixture was incubated at 68°C for 10 min. 2 µl of RiboLock was added and the mixture was incubated at 37°C for 30 min. Following the incubation the RNA/probe mix was added to Round 1 beads (2xB&W) and incubated for 15 min at 37°C, 1000 rpm. The beads were pelleted on a magnetic stand and the supernatant was transferred to Round 2 beads (1xB&W) and incubated for 15 min at 37°C, 1000 rpm. The supernatant was collected and purified with Monarch RNA Cleanup Kit. Six 3' biotinylated probes were used for ribodepletion. Five for LSU and one for SSU. Ribodepletion probes are listed in Supplementary Table 2.

ONT library preparation and sequencing

Three µg of input RNA (either native mitochondrial RNA, including bisulfite converted and FTO-treated samples or RNA obtained via in vitro transcription) was polyadenylated with Poly(A) Polymerase Tailing Kit (Bioresearch Technologies) and purified with RNAClean XP beads (Beckman Coulter). Resulting polyadenylated RNA was used for library preparation with SQK-RNA002 Direct RNA Sequencing Kit (ONT). The reverse transcription step was omitted [24]. RNA libraries were sequenced on MinION sequencer with FLO-MIN106D Spot on R9 flow cells for 24–36 h.

Data analysis pipeline

Basecalling and read mapping

Preprocessing of raw ONT data, including basecalling, was performed with MoP2 preprocess module [25]. The basecaller used was Guppy version 3.6.0 with default settings (quality threshold of 5 within the MoP2 pipeline) in HAC model [26]. Basecalled reads from two sequencing runs for each strain were merged and then mapped to mitochondrial genome [27]. As the *C. albicans* mitochondrial genome contains two identical, inverted repeat regions [28, 29], the second of these regions (IRb) was removed from the reference used in this study in order to maintain the directionality of transcript reads [30]. Numbers of reads from each run and merged reads mapping are shown in Supplementary Table 5. The unique characteristics of the *C. albicans* mitochondrial genome, with its relatively small size and high AT content, necessitated the selection of an appropriate read mapping strategy.

Three available programs were tested: minimap2 [31], GraphMap [32], and BWA [33]. The most comprehensive view of the transcriptome was achieved by combining the results from minimap2 in transcript assembly mode, aided by 2passtools [34] (with a reference containing intron coordinates) and short read mapping with either minimap2 (minimap2 -ax map-ont -k8 -w1 -m30 -s30 -g40 -G40) or BWA (bwa mem -W13 -k6 -xont2d -T20). Initial mapping was performed in transcript assembly mode with minimap2/2passtools, for mapping of long reads, including those corresponding to intron-containing genes. Unmapped reads were then extracted with samtools [35]. This step was taken to prevent multiple mapping of the same reads. Read that were not mapped in the previous step served as input for mapping in short read mapping mode (either with minimap2 or BWA). BWA short read mapping provided the best coverage for tRNA transcripts and was thus used in subsequent analyses (Supplementary Table 7 and Supplementary Fig. 5). It should be noted however, that minimap2 with shorter k-mer parameter (-k8 instead of otherwise recommended k-15) also provided better tRNA mapping results than previously reported for this mapper with *S. cerevisiae* cytoplasmatic tRNAs [36]. This result might be an indication, that *C. albicans* mitochondrial tRNAs are not as heavily modified as their cytoplasmatic counterparts in either yeast species. The resulting bam files for long and short reads were subsequently merged. The mapping results were visualized with pyGenomeTracks [37]. For a simplified view of transcript species present in the transcriptome, RATTLE analysis for visualizing consensus reads was also performed [38]. The scripts used in the read mapping procedure along genome sequence, annotation and intron coordinates files are available at repod.icm.edu.pl.

Isoform analysis of intron-containing transcripts

The results from mapping with minimap2 in transcript assembly mode (prior to merging with short read mapping results) for the BWP17 strain as well as ΔCaDSS1 and ΔCaPET127 deletion strains served as input for isoform analysis. LIQA [39] was used for quantifying levels of discrete transcript isoforms for three intron-containing genes: rnl, COB and COX1. The commands and isoform annotation files used in this analysis are available at repod.icm.edu.pl.

Modification detection, identification and quantification

To identify modified positions, sequencing results from each of the three strains were compared with those from in vitro-transcribed mitochondrial transcripts of *C. albicans*. Modification detection was carried out using the MoP2 package [25], which incorporates Nanopolish [40], Tombo [41], EpiNano [9], and Nanocompore [42] programs. Due to relatively short read lengths and probable high density of modifications, tRNAs, couldn't be reliably analyzed with MoP2 pipeline as individual transcripts. Modifications could however be detected when tRNAs were analyzed as a part of longer transcription units, including, but not limited to the part of TU5 (along with NAD2 and NAD3) or COX2 pre-mRNA. Individual tRNAs were also analyzed with NanoRSM [43] (with paired condition pseudouridine detection script) and ELIGOS (pair_diff_mod) [44]. Both of these tools take advantage of the fact that some RNA modifications result in erroneous basecalling (also known as miscalling) in ONT direct RNA sequencing [42, 45, 46]. The most prominent type of miscalls in ONT data is U to C, indicative of pseudouridine modification. Detection of miscall-associated modifications other than pseudouridines was conducted with ELIGOS. Positions with coverage over 100, $p < 0{,}005$ and the ratio of an indicative miscall greater than 10% were selected. Separate analysis for pseudouridine quantification was performed for mature tRNAs and pre-tRNAs. Reads were filtered by lengths with transcripts below 70 nt classified as mature and over 140 nt as pre-tRNA (transcripts of intermediate lengths were excluded from the analysis). Sequencing miscalls have been visualized with Golden Helix GenomeBrowser 3.1.0 [47]. m6A detection was performed with CHEUI (Methylation (CH3) Estimation Using Ionic current) [48] with FTO treated/untreated WT native RNA and in vitro transcribed RNA sequencing data as inputs. FTOvsIVT output was filtered by removing all entries such that: stoichiometry ratios were below 0.8 or above 1.2, statistics parameter was below -2 or above 2 or p-value was below 0.1. This was intended to remove all k-mers where a significant difference was detected between the FTO-treated sample and IVT sample (as the intersection between unfiltered WTvsFTO and WTvsIVT sets excluding the unfiltered FTOvsIVT set yielded no entries). The filtering results were intersected with WTvsFTO and WTvsIVT outputs. The output was then filtered by keeping entries with p-value under 0.005, stoichiometry ratio above 1.5; statistics parameter greater than 3, a positive stoichiometry difference and coverage > 100 (for all listed parameters in both WTvsFTO and WTvsIVT comparisons). The bam files from MoP2 preprocessing module were used as inputs for both CHEUI and MoP2 modification detection.

ELIGOS, NanoRSM and CHEUI were all run on default settings. MoP2 configuration and reference files, along with the ELIGOS, NanoRSM and CHEUI commands are made available at repod.icm.edu.pl.

Bisulfite sequencing analysis

Fastq output files were mapped to the reference sequence where all cytosines were substituted for thymines (either in normal or inverted orientation, the latter for reverse strand transcripts). Due to RNA fragmentation resulting from bisulfite treatment as well as multiple spurious alignments observed when BWA short read mapping (bwa mem -W13 -k6 -xont2d -T20) was employed, two different mapping strategies were used. These involved either three rounds of mapping (long spliced read mapping, followed by short read mapping with minimap2 and BWA mapping with standard settings) or short read mapping with minimap2. T-C mismatches were detected and quantified with ELIGOS and NanoRSM separately for each type of mapping, with synthetic converted RNA serving as control. With ELIGOS analysis, positions detected in native RNA (compared with the in vitro transcribed control) where C was the major allele, and where C counts were higher in the native sample than in the synthetic sample were selected. For NanoRSM analysis, positions where the C ratio between native and in vitro transcribed RNA was at least 3 were selected. A minimum coverage of 20 per position was set for each analysis. Scripts for bisulfite read mapping and ELIGOS/NanoRSM analysis are available at repod.icm.edu.pl.

Results

Read distribution and coverage

A read distribution comparable to the one achieved with Ion Torrent sequencing was obtained for all mitochondrial protein-coding transcripts and rRNAs [30], however the coverage for individual exons was significantly more uniform, indicating that the drops in coverage observed for exonic regions in Ion Torrent sequencing were in fact artifacts resulting from the library preparation process (Fig. 1). There was also notably higher relative coverage in intragenic regions of mitochondrial transcription units. The majority of reads were mapped to the large subunit rRNA (74,17% compared to 61.23% in Ion Torrent RNA-seq data). The relative abundance of reads mapped to the small subunit rRNA (rns), as well as reads for all coding transcripts was lower for ONT sequencing (Table 1 and Fig. 2). This is primarily due to increased representation of reads corresponding to tRNAs, demonstrating that two-stage mapping, along with in vitro polyadenylation, which allows for overcoming the 200bp threshold of MinION software (Supplementary Fig. 1), more than compensates for the previously reported issue of insufficient tRNA coverage [11] and provides a straightforward method for obtaining a comprehensive view of organellar tRNAs and other short transcripts. However, it also underscores the degree to which the mapping strategy affects all the subsequent stages of transcriptome analysis.

Transcription units and transcript ends

The *Candida albicans* mitochondrial genome contains 8 policistronic transcription units (TUs) [30]. ONT sequencing enabled (both through direct observation and generation of consensus reads by RATTLE) the detection of long reads covering multiple genes within these transcripts, such as NAD6-NAD1 in TU01, ATP8-ATP6 in TU04, NAD2-NAD3 in TU05 and NAD4L-NAD5 in TU07, as well as multiple cases of reads covering coding transcripts along with adjacent tRNAs (Fig. 3A). Additionally, it was found that TU02 is considerably longer at the 3' end than previously reported [30], reaching as far as the end of an unidentified transcript at position 8940. The 5' and 3' ends observed for some mature transcripts have also been found to differ from those reported previously (Fig. 3B) [30]. Most of said changes, summarized in Table 2 are relatively small and may be due to artifacts from library preparation procedure for Ion Torrent sequencing. A significant 3' extension of more than 100nt was also found for TU03, however its heterogeneity makes it impossible to define a specific 3' end position. Of particular interest are the 5' ends of TU05 and TU06, which have been found to be 8 and 6 nucleotides downstream of putative transcription start sites, respectively [30]. Most reads are in fact an additional 2–3 nucleotides shorter, and no reads corresponding in length to putative transcription start sites have been detected. If this is a consequence of posttranscriptional processing, it would not involve exonucleolytic truncation by CaPet127, as 5' ends corresponding to the predicted promoter + 1 sites were not detected in ΔCaPET127 strain either. It should be noted however, that 5' end mapping by direct ONT direct RNA sequencing is also prone to artifacts, primarily 3' coverage bias, unless accompanied by 5' adapter ligation [49, 50]. Thus, these relatively small changes at the 5' ends will require further verification.

Qualitative differences were also observed between the WT and deletion strains (Fig. 4). CaPet127 is a 5'→3' exonuclease — thus, changes at the 5' ends of transcripts were to be expected in ΔCaPET127 strain. This was indeed found for the TU08 and TU06/COB transcripts, where discrete isoforms with 5' extensions can be detected (Fig. 4A). These elongated transcripts are present along with properly truncated ones. This was to be expected for TU08, as the downstream tRNA genes would be excised from the transcript during the process of maturation. However, 5' elongation is also observed for TU06/COB, although at a relatively low level of approximately 3% (compared to TU08, where at least 40% of reads have 5' elongations). This could suggest the

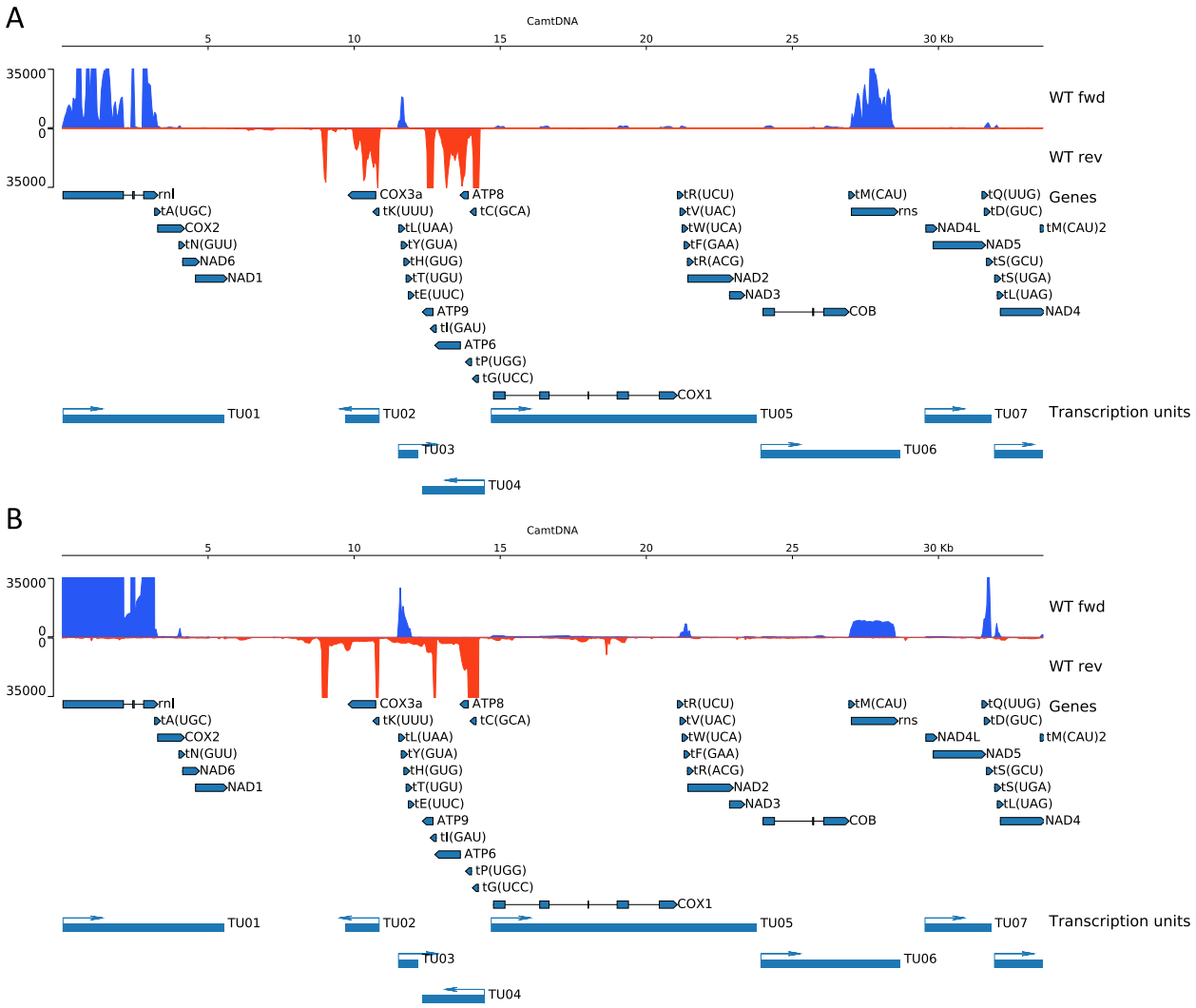

Fig. 1 Mitochondrial RNA from wild type *C. albicans* (strain BWP17) mapped to mtDNA. Forward strand coverage is marked in blue. Reverse strand coverage is marked in red; **A** – reads from Ion Proton platform; **B** – reads from ONT MinION platform The image was generated with pyGenomeTracks [37]

presence of alternative transcription start sites which produce longer transcript that are subsequently truncated by CaPet127. However, a more parsimonious and indeed more likely explanation for the observed TU08 5' extension might be that it is co-transcribed with TU07 (as transcripts overlapping both TUs were also observed). Thus what was previously believed to be two separate transcription units might in large part originate from single primary transcript, with endonucleolytic cleavage at the 3' end of tS(GCU)in early stages of maturation producing two secondary transcripts. One might speculate if such an expression system, where TU08 might be transcribed both as a part of longer transcript along with TU07 and from its own promoter contributes to regulatory the flexibility of *C. albicans* mitochondria. Likewise, for TU06/COB the 5' extensions might be a result of a low level readthroughs from TU5. However, in this case any physiological significance of such events is far less likely.

The ΔCaDSS1 strain displayed a stronger phenotype and thus a higher level of intermediate transcripts corresponding to intragenic regions (also outside of established transcription units) was observed, which is in line with previously published results [13]. Since CaDSS1 is a 3'→5' exoribonuclease, long readthroughs were also observed (like in TU02, TU03 and TU04) (Fig. 4B). A short transcript of approximately 300nt was also detected at the 3' end of TU06, directly downstream

Table 1 Relative abundances of reads corresponding to mitochondrial genes (excluding rnl) obtained from ONT sequencing, mapped in a two-step fashion (with minimap2 and BWA) and from Ion Torrent sequencing. Genes with a relative increase in read number in ONT sequencing are marked in blue. Genes with decreased relative read number are marked in red

Gene ID	WT ONT minimap2+bwa	WT Ion Torrent
tA(UGC)	3.98%	2.61%
COX2	0.22%	0.99%
tN(GUU)	0.00%	0.00%
NAD6	0.04%	0.17%
NAD1	0.04%	0.44%
COX3a	0.14%	0.85%
tK(UUU)	1.70%	0.23%
tL(UAA)	7.75%	1.26%
tY(GUA)	9.70%	4.92%
tH(GUG)	10.89%	1.02%
tT(UGU)	3.23%	0.24%
tE(UUC)	4.65%	0.05%
ATP9	0.10%	1.37%
tI(GAU)	1.40%	0.02%
ATP6	0.04%	0.89%
ATP8	0.03%	0.32%
tP(UGG)	2.09%	0.05%
tC(GCA)	1.77%	0.91%
tG(UCC)	4.57%	0.74%
COX1	1.30%	4.72%
tR(UCU)	0.19%	0.04%
tV(UAC)	3.26%	0.30%
tW(UCA)	1.45%	0.19%
tF(GAA)	7.87%	0.17%
tR(ACG)	0.00%	0.00%
NAD2	0.13%	0.49%
NAD3	0.02%	0.11%
COB	0.56%	2.99%
tM(CAU)	1.84%	0.06%
rns	4.43%	70.41%
NAD4L	0.03%	0.05%
NAD5	0.11%	0.94%
tQ(UUG)	4.06%	0.01%
tD(GUC)	6.90%	0.43%
tS(GCU)	11.71%	0.85%
tS(UGA)	3.23%	0.47%
tL(UAG)	0.41%	0.02%
NAD4	0.07%	0.66%
tM(CAU)2	0.08%	0.00%

of rns. It is present both in WT and deletion strains; however, in deletion strains, reads covering both said transcript and rns were also detected (Fig. 4C). This suggests that it is a product of a readthrough from rns, that is subsequently endonucleolytically excised (the presence of unexcised transcripts in deletion strains being a secondary effect of mitochondrial dysfunction).

Isoform analysis of intron-containing transcripts

The *C. albicans* mitochondrial genome includes three intron-containing genes: rnl, COB and COX1. Unlike rnl, where mature rRNA makes up the majority of transcript species, for COB and COX1 the levels of mature isoforms appear to be underestimated. These results might be a consequence of lower coverage of these transcripts and in the case of COX1 also of transcript complexity and a very short exon 3 (of mere 10nt), and thus all the results from this section should be treated as semiquantitative. Relatively high levels of three different partially spliced isoforms of the COX1 transcript, missing either intron 1 (19.27%), intron 4 (10.88%) or intron 3 (5.33%) were detected. The percentages of COB transcripts retaining either intron 1 (7.47%) or intron 2 (9.08%) were also comparable. This finding indicates that splicing of these transcripts is not strictly a sequential process. For rnl on the other hand, the abundance of transcripts retaining intron 2 was nearly 40× higher than of those retaining intron 1, implying a considerable degree of sequentiality.

Changes in the relative abundances of transcript isoforms were observed between the WT strain and deletion strains. A decrease in mature transcript levels and an increase in unspliced transcript levels were very pronounced in ΔCaDSS1 strain, with ΔCaPET127 exhibiting intermediate levels of mature and unspliced transcripts between those of the WT and ΔCaDSS1 strains. This result is in line with previous studies indicating that ΔCaDSS1 has considerably more severe mitochondrial defects, as well as Northern blot, NGS and reverse transcription analyses for ΔCaDSS1 and ΔCaPET127 indicating, respectively: lower levels of mature mitochondrial transcripts, higher intron coverage and an increase in retention [13, 14]. Changes in levels of spliced "intermediate" isoforms were also observed for each gene (Table 3). These findings are broadly consistent in deletion strains when compared to WT strain, indicating that splicing impairment is mainly a secondary effect of general mitochondrial dysfunction. However, some differences between the deletion strains were also observed. The relative abundance of isoforms retaining intron 2 in COB and introns 3 and 4 in COX1 was decreased in ΔCaPET127 but increased in ΔCaDSS1, and COB intron 1 and COX1 introns 1 and 2 were retained at higher relative levels in ΔCaPET127, which suggests that 5'/3' exonucleolytic RNA processing might affect the splicing of adjacent introns.

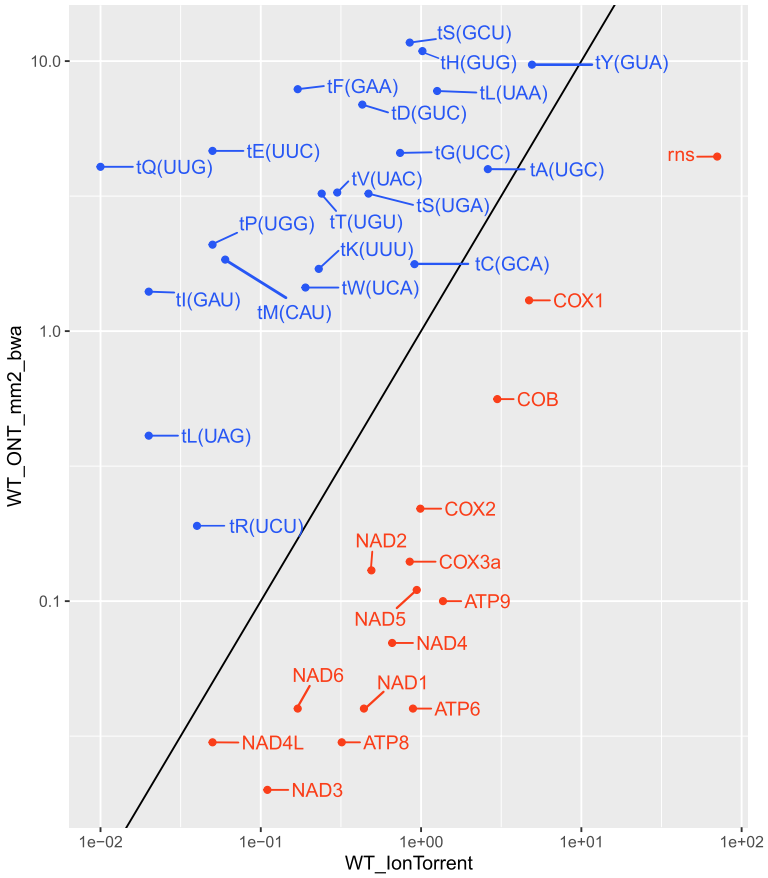

Fig. 2 Scatter plot of relative abundances of reads corresponding to mitochondrial genes (excluding rnl) obtained from ONT and Ion Torrent sequencing. Reads corresponding to genes above the diagonal line are more abundant in ONT data, and reads corresponding to genes below the line are more abundant in Ion Torrent data. Color coding convention is the same as in Table 1

Modification detection, identification and quantification

Comparing sequencing data from native RNA (be it from WT or deletion strains) with in vitro transcribed (IVT) RNA of the same sequence allows for the detection of putative posttranscriptional modification sites (as in vitro transcribed RNA contains no such modifications). This was conducted for *C. albicans* mitochondrial transcriptome via the MoP2 pipeline, with IVT RNAs covering all mitochondrial coding sequences, rRNAs and tRNAs serving as unmodified reference. Due to relatively low coverage for most coding transcripts, MoP2 analysis reliably indicated putative modification sites only for rnl, rns and COX2 transcripts (Fig. 5A). 1 out of 8 detected sites, found in rnl, corresponded to a splice junction. This might indicate the presence of a modification involved in or resulting from RNA splicing, but may also be an artifact of misalignment during the mapping step. However, all the sites mentioned above are detected repeatedly and for a wide range of mapping parameters. Additionally, two putative modification sites found towards the 3' end of COX2 transcript correspond to Asn tRNA, which is a part of COX2 pre-mRNA. No qualitative differences were detected between the WT strain and deletion strands in regard to modifications described above.

Further analyses were conducted on partially ribodepleted WT RNA, which reduced the percentage of rRNA reads from 96–98% to 61–66% (Supplementary Table 6). Here, a number of individual putative modification sites were detected on coding transcripts (Fig. 5B). In the case of COX1 transcripts, all sites found were located in or around positions corresponding to splice sites. A considerable number of said sites are found towards the 3' ends of transcripts. For NAD2 and NAD4 modifications were detected for cotranscribed tRNAs, respectively at the 5' and 3' ends of transcripts. Additionally, when entire transcription units (or large parts thereof) are being analyzed, clusters of modifications corresponding to tRNAs are detected (Fig. 5C), hinting at a high stoichiometric level of modifications in tRNAs relative to coding transcripts (MoP2 pipeline is not conducive for analyzing individual short and densely modified transcripts, such

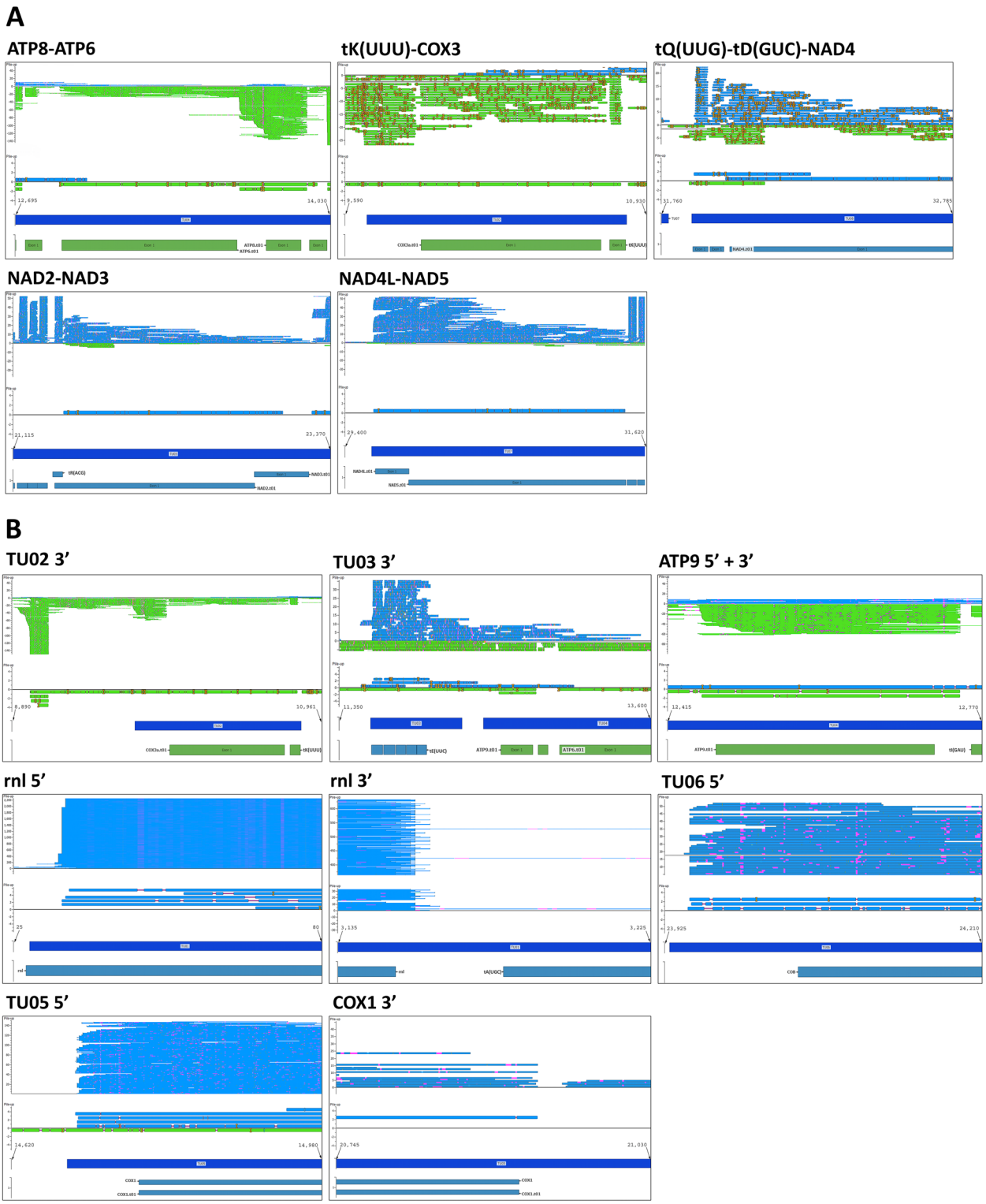

Fig. 3 A visualization of ONT reads corresponding polycistronic transcripts (**A**) and 5′/3′ ends that differ from the ones previously obtained via Ion Torrent sequencing (**B**). Top track represents raw reads. Below it are consensus isoforms obtained with RATTLE. Images generated by GoldenHelix Genome Browser [51]

Table 2 Transcript ends found to be different in ONT sequencing (changes relative to Ion Torrent annotations are given in brackets)

Transcript	Strand	End positions on mtDNA	
rnl	+	45(+10)	3153(+4)
TU02	-	8940(-743)	10,854
ATP9	-	12,726(-257)	12,726(+27)
ATP6	-	12,883(-12)	13,635
TU05	+	14,701(+8)	23,769
COX1	+	14,771	20,936(+10)
TU06	+	23,933(+6)	26,808

as tRNAs). This analysis yielded a general image of the modification landscape for mitochondrial transcripts in terms of modified positions. However, with the notable exception of pseudouridines (further elaborated on later in this section), it does not provide information on specific modification types. Attempts to identify three types of common RNA modifications, m6A (N6-methyladenosine), Ψ (pseudouridine) and m5C (5-methylcytidine) are described below.

FTO is a m6A RNA demethylase. Thus, a comparison between native WT RNA, FTO-treated WT RNA and in vitro transcribed RNA should enable the identification of m6A sites in mitochondrial transcripts. Data from FTO-treated RNA was analyzed with Nanopolish-based CHEUI pipeline [48], with untreated WT RNA and in vitro transcribed RNA serving as controls. Eight putative m6A positions were detected following a three-way comparison and filtering (see *Materials and Methods*). However, only one, detected in a k-mer 303–312 of rnl corresponded to a k-mer detected by MoP2 comparison of WT and FTO treated samples. This k-mer also had the smallest detected stoichiometry difference in FTOvsIVT comparison among filtered sites and nearly identical results for WTvsIVT and WTvsFTO comparisons. K-mer 1395–1404 in rns did correspond to the highest scoring raw k-mer in the MoP2 analysis, even though no k-mer in WT/FTO comparison in MoP2 analysis passed the pipeline's filtering. For COB k-mer 855 a similar situation was observed, with an overlapping MoP2 k-mer scored 4th on the raw k-mer list (Table 4 and Supplementary Tables 3 and 4). A subsequent LC–MS and dot blot analyses found relatively low levels of m6A in untreated WT sample (Supplementary Figs. 2, 3 and 4). This suggests a small number of m6A positions in *C. albicans* mitochondrial transcriptome and/or low m6A methylation level. This also makes the estimation of FTO-mediated demethylation efficiency problematic. Thus, the putative m6A sites listed in Table 3 will require independent experimental verification.

A separate analysis of tRNA transcripts revealed a high level of U-to-C miscalling, indicative of the presence pseudouridine (Ψ) [43, 52] (Table 5). Majority of uracils in position ~ 54 wa miscalled as cytosines indicating the presence of pseudouridines in respective reads, both in pre-tRNAs and mature tRNAs, which suggests that the modification takes place prior to tRNA excision from parent transcript. Other uracils, mainly in positions 25–41 were also modified, but to a lesser extent (Fig. 6). The miscall levels for these positions were also much lower for pre-tRNAs than for mature tRNAs, indicating that they are primarily modified following tRNA excision (Fig. 7). These results are in line with results obtained in vitro for bacterial tRNAs [53].

Low levels (of approximately 20%) of A-to-G and G-C miscalls, indicative of inosine [54] and m7G [45] modifications respectively, have also been detected in a number tRNA sites (Fig. 6, Supplementary Table 8). However, the relatively low level of these miscalls along with lower repeatability within and among tRNA transcripts means that these results warrant further experimental verification.

m5C modifications can be detected via bisulfite sequencing, wherein unmodified cytosines are converted to uracils, while m5C positions are not. Bisulfite conversion of mitochondrial RNA has previously been reported to occur with relatively low efficiency [55]. This was also the case with *Candida albicans* mtRNA, which necessitated the use of in vitro transcribed RNA as control and limited the analysis only to positions with high levels of m5C methylation. Bisulfite sequencing also served as an additional control for pseudouridine detection, as U-to-C miscalls due to Ψ modification are not affected by bisulfite conversion. Putative pseudouridine modifications of tRNAs at position ~ 54 were in fact the ones with strongest signals in ELIGOS and NanoRSM analyses. Due to fragmentation occurring during the bisulfite conversion, as well as spurious alignments that necessitated the exclusion of BWA short read mapping, both native and synthetic RNA were mapped either with minimap2/BWA (with long read minimap2 mapping and subsequent mapping of short reads with minimap2 and BWA with standard settings) or with short read mapping with minimap2. The list of positions thus detected is shown in Table 6. Three putative m5C sites were detected: in tyrosine, glutamine and serine tRNAs. The site in tY(GUA) was detected with high confidence both with standard and short read mapping, for both ELIGOS and NanoRSM analyses. The tS(GCU) site was detected by both algorithms with short read mapping and by ELIGOS with standard mapping. The tQ(UUG) site was detected only by NanoRSM with short read mapping.

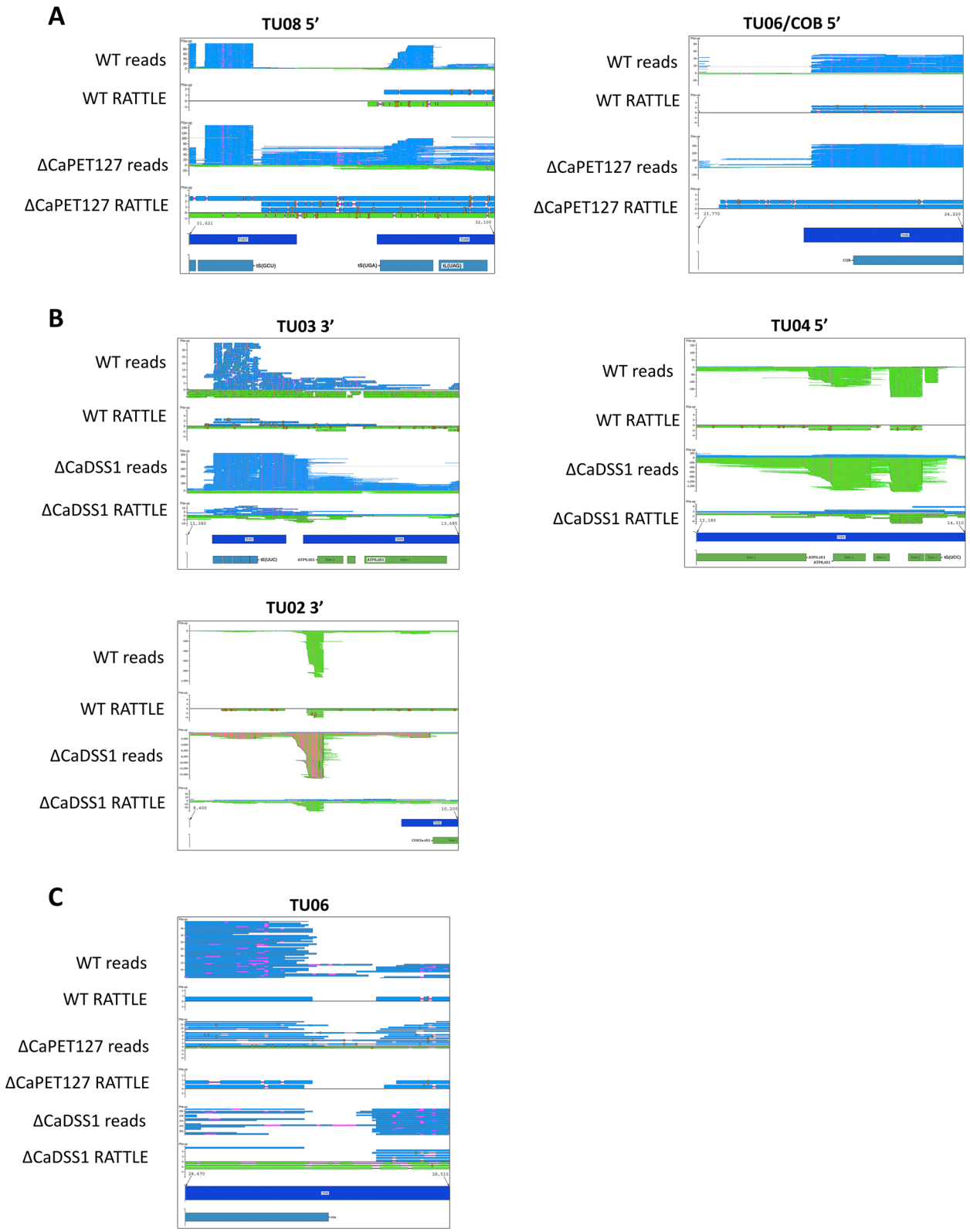

Fig. 4 Qualitative differences observed in transcript ends between the WT and deletion strains of *C. albicans*. A pileup of individual reads and consensus RATTLE reads is shown for each strain. **A**—changes observed for ΔCaPET127 strain; **B**—changes observed for ΔCaDSS1 strain; **C**—a non-coding transcript found at 3' end of TU6

Table 3 Relative abundances (calculated as percentages of all reads assigned to a given transcript) of mitochondrial transcript isoforms in WT *C. albicans* and differences in abundances between WT and deletion strains as well as differences between the deletion strains

isoform	WT	ΔCaPET127vsWT	ΔCaDSS1vsWT	ΔCaDSS1vsΔCaPET127
rnl_mature	**71,97%**	↓x1,02	↓x1,12	↓x1,09
rnl_unspliced	0,54%	↑x1,16	↑x5,02	↑x4,32
rnl_unsp_i1	0,70%	↑x1,02	↑x3,56	↑x3,49
rnl_unsp_i2	26,78%	↑x1,05	↑x1,13	↑x1,07
COB_mature	**20,17%**	↓x2,72	↓x100,92	↓x37,13
COB_unspliced	63,28%	↑x1,28	↑x1,4	↑x1,1
COB_unsp_i1	7,47%	↓x1,71	↓x6,76	↓x3,96
COB_unsp_i2	9,08%	↓x1,26	↑x1,09	↑x1,36
COX1_mature	**5,34%**	↓↓	↓↓	ND
COX1_unspliced	56,26%	↑x1,03	↑x1,31	↑x1,27
COX1_unsp_i1	ND	ND	ND	ND
COX1_unsp_i2	ND	ND	ND	ND
COX1_unsp_i3	ND	↑↑	ND	↓↓
COX1_unsp_i4	ND	ND	ND	ND
COX1_unsp_i12	ND	↑↑	↑↑	↓x20,1
COX1_unsp_i13	ND	ND	ND	ND
COX1_unsp_i23	0,02%	↑x311,67	↑x12,37	↓x25,19
COX1_unsp_i34	1,12%	↓x1,28	↑x2,41	↑x3,08
COX1_unsp_i14	ND	ND	ND	ND
COX1_unsp_i24	ND	↑↑	ND	↓↓
COX1_unsp_i123	10,88%	↓x1,99	↓x7,91	↓x3,97
COX1_unsp_i234	19,27%	↑x1,11	↓x1,48	↓x1,64
COX1_unsp_i134	ND	↑↑	↑↑	↑x2,18
COX1_unsp_i124	5,33%	↓x4260,19	↓x1,16	↑x3672,4

↑↑/ ↓↓ indicate a change exceeding 3 orders of magnitude (usually with very low abundance in one of the samples); ND – not detectable or below 1E-5 in both samples; relative increase in abundance is additionally indicated in hues of blue, and relative decrease in hues of red. Supplementary Table 9 contains corresponding data in raw form

Discussion

In this study, we demonstrated the viability of applying direct RNA sequencing to the mitochondrial transcriptome of *C. albicans* and by extension, other yeast species. By using a tailored approach for read mapping, we obtained sufficient coverage for both long and short reads generated by ONT sequencing, at least partly compensating for issues reported with short-read mapping [11]. Analysis of transcript isoforms in the wild-type, ΔCaDSS and ΔCaPET127 strains has demonstrated the utility of direct RNA sequencing for 5' and 3' end mapping and primary transcription unit detection, as well as for quantifying changes in mitochondrial splicing, providing a straightforward method of investigating the effects of deletions/mutations in proteins involved in posttranscriptional processing in mitochondria. The relative abundances of transcript isoforms retaining certain introns have been found to be higher either in ΔCaPET127 or ΔCaDSS1 strain, which suggests a connection between 5'/3' exonucleolytic RNA processing and the splicing of adjacent introns. Still, it is worth noting, that the pleiotropic character of mitochondrial

(See figure on next page.)

Fig. 5 Putative modification sites detected by NanoConsensus tool from Master of Pores 2 pipeline. In each case reads from wild type *C. albicans* were compared against unmodified RNA obtained via in vitro transcription. Sites corresponding to splice junctions are marked with asterisks. tRNAs and tRNA clusters are marked with red boxes. **A**—Results for rnl, rns and COX2 transcripts obtained from sample without ribodepletion. **B**—Results for coding transcripts obtained from partially ribodepleted sample. **C**—Results for *C. albicans* transcription units TU4, TU7, TU8 and a part of TU5 encompassing a tRNA cluster, NAD2 and NAD3

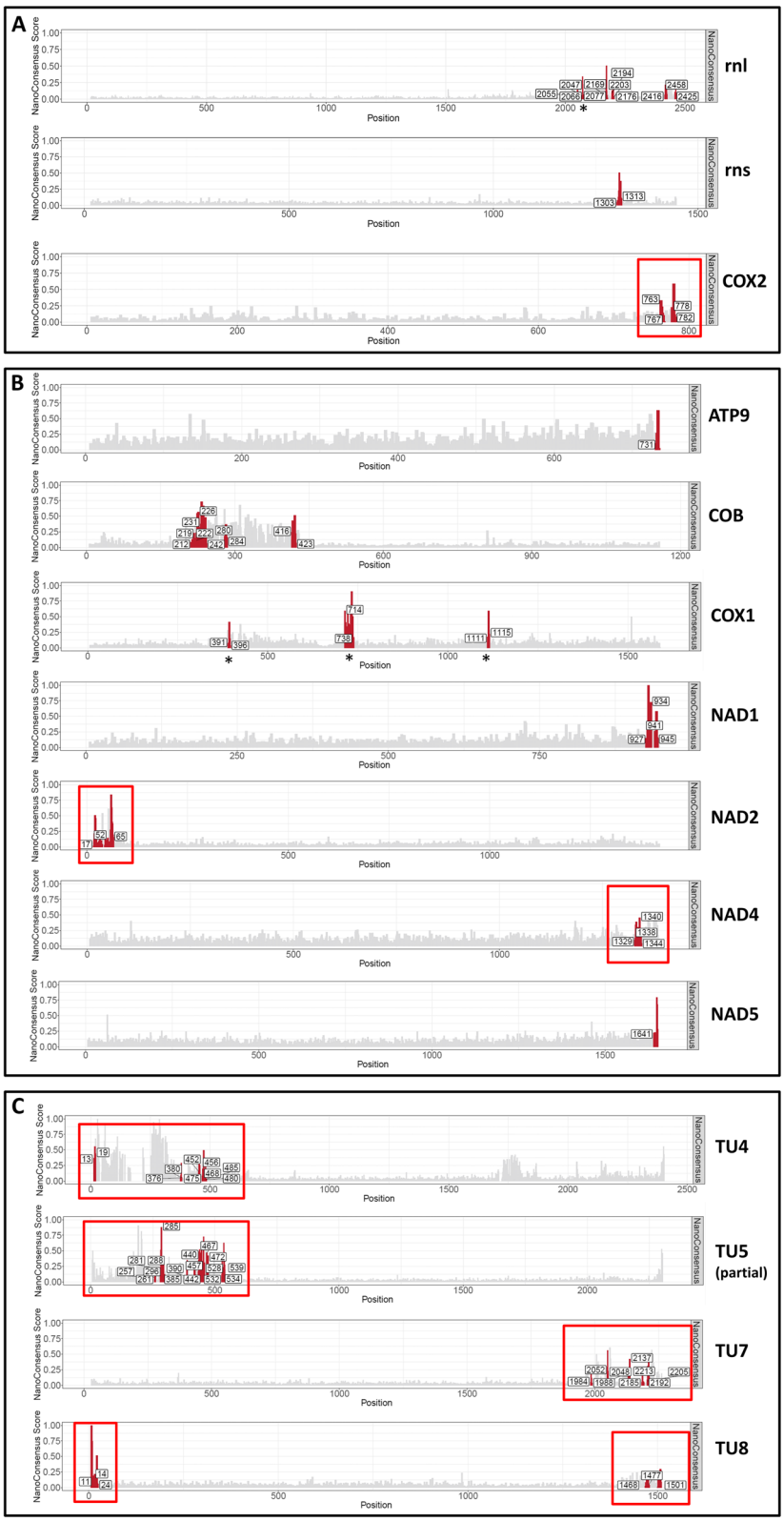

Fig. 5 (See legend on previous page.)

Table 4 Putative m6A sites detected by Nanopolish/CHEUI. K-mer overlapping with independently detected, supported MoP2 k-mer is marked in blue. K-mers overlapping with high-scoring raw MoP2 k-mers are marked in blue

ID	p value	stoichiometry ratio
COB_581_ATGTATGCA	8.84E-05	1.761562998
COB_855_CCTGATAAA	0.000522015	2.063157895
COX1_1144_TATTATCAA	4.20E-05	1.895604396
COX1_1417_GTGTATTAG	0.000287561	2.08
rnl_1430_TGATACCTG	5.20E-63	1.651502971
rnl_2165_TAGGACTGT	1.00E-77	2.192885987
rnl_303_TGTTATTGA	6.58E-82	1.823311937
rns_1395_TGAAATACA	3.19E-33	1.802193103

dysfunctions makes it difficult to draw a clear line between the specific effects of 5'/3' processing (or lack thereof) and the secondary effects that the deletions of CaPET127 and CaDSS1 have on splicing. However, one might cautiously speculate that intermediate splicing isoforms accumulated to a higher degree in ΔCaPET127, compared both to WT and ΔCaDSS1 strains, are indeed directly related to 5' processing, as the consequences of CaPET127 deletion (both on phenotypic and molecular levels) are otherwise relatively mild.

These results provide a new insight *C. albicans* mitochondrial transcriptome and could not have been achieved with short read NGS. It should be noted however, that they are in line with previously conducted studies. ONT data presented in this study has indeed confirmed mechanisms that have previously been hinted at, such as the presence of additional promoters within regions corresponding to larger transcription units [30], as well as the fact that mitochondrial exonucleases play a crucial role in molding the mitochondrial transcriptome by concealing pervasive transcription [13]. Results pertaining to transcript ends and transcriptional readthroughs that haven't been previously observed or contradict the findings from Ion Torrent sequencing will require further verification. Tailored RT-PCR experiments, such as circularized RT-PCR [56] would enable the confirmation of said results.

Attempts to detect and identify transcript modifications yielded promising results, however the methodology involved requires further development. It should be noted, that RNA modification data for *C. albicans* is extremely sparse, and limited to cytosolic leucine tRNA, according to Modomics [57]. To the authors' best knowledge, this is the first work reporting any RNA modifications in *C. albicans* mitochondrial transcriptome. Unsurprisingly, mitochondrial tRNAs appear to be the most densely modified transcripts, with easily detectable pseudouridine and m5C sites. Miscall analysis also allowed us to detect a number of putative inosine and m7C sites. Additionally, we were able to determine the stage at which the introduction of pseudouridines at different positions occurs during tRNA maturation by simply applying size cutoffs for ONT reads. Pseudouridine modifications serve primarily to ensure structural stability of tRNA [58]. m5C promotes base pairing, translation, and is involved in stress response [59]. m7G is involved in tRNA biosynthesis and stabilization [60]. The primary function of inosine is the expansion of codon recognition when present at the first anticodon position [61]. Puzzlingly, none of the inosine modifications detected in this study were at that position, with most being present either at the third anticodon position or just behind the anticodon sequence. Pseudouridine, m5C and inosine have also been found to serve as identity elements for aminoacylation [62]. Aberrations in all the types of tRNA modifications detected in this study have also been linked to multiple human diseases [63]. Determining the significance of each tRNA modification in *C. albicans* will require further investigation, possibly involving comparative analysis of mitochondrial tRNA modifications across different yeast species. It should be noted that the list of tRNA m5C sites detected in this study is almost certainly not exhaustive (neither in regard to mitochondrial tRNAs, nor the mitochondrial transcriptome as a whole), due to a combination of incomplete bisulfite conversion – which necessitated strict filtering of putative sites – and insufficient coverage of some mitochondrial transcripts. Likewise, only pseudouridine sites with sufficient coverage were included in the list of modified positions detected via bisulfate sequencing, thus the absence of some previously detected modifications should not cast doubt on their presence (as all tRNA pseudouridine sites at position ~ 54 were detected whenever the coverage was sufficient).

Modifications of rRNAs and coding transcripts are considerably more challenging to identify than those within

Table 5 Putative pseudouridine (Ψ) sites in tRNA transcripts. Sites corresponding to position ~ 54 in their respective transcripts are marked in bold. CamtDNA postion signifies the position on *Candida albicans* mitochondrial genome corresponding modification sites in tRNA transcripts

CamtDNA position	tRNA transcript	Strand	CamtDNA position	tRNA transcript	Strand
3208	tA(UGC)	+	14,143	tC(GCA)	-
3224	**tA(UGC)**	**+**	14,144	tC(GCA)	-
4054	**tN(GUU)**	**+**	**14,199**	**tG(UCC)**	**-**
10,796	**tK(UUU)**	**-**	14,226	tG(UCC)	-
10,814	tK(UUU)	-	21,097	tR(UCU)	+
10,815	tK(UUU)	-	21,098	tR(UCU)	+
10,816	tK(UUU)	-	21,103	tR(UCU)	+
10,817	tK(UUU)	-	21,105	tR(UCU)	+
11,558	tL(UAA)	+	**21,118**	**tR(UCU)**	**+**
11,559	tL(UAA)	+	21,180	tV(UAC)	+
11,564	tL(UAA)	+	21,193	tV(UAC)	+
11,567	tL(UAA)	+	**21,207**	**tV(UAC)**	**+**
11,588	**tL(UAA)**	**+**	21,259	tW(UCA)	+
11,647	tY(GUA)	+	21,260	tW(UCA)	+
11,648	tY(GUA)	+	**21,280**	**tW(UCA)**	**+**
11,650	tY(GUA)	+	21,329	tF(GAA)	+
11,654	tY(GUA)	+	21,337	tF(GAA)	+
11,664	tY(GUA)	+	**21,351**	**tF(GAA)**	**+**
11,678	**tY(GUA)**	**+**	21,445	tR(ACG)	+
11,749	tH(GUG)	+	**21,460**	**tR(ACG)**	**+**
11,756	**tH(GUG)**	**+**	**26,987**	**tM(CAU)**	**+**
11,806	tT(UGU)	+	31,525	tQ(UUG)	+
11,810	tT(UGU)	+	31,526	tQ(UUG)	+
11,834	**tT(UGU)**	**+**	31,531	tQ(UUG)	+
11,886	tE(UUC)	+	**31,545**	**tQ(UUG)**	**+**
11,888	tE(UUC)	+	**31,623**	**tD(GUC)**	**+**
11,889	tE(UUC)	+	31,679	tS(GCU)	+
11,892	tE(UUC)	+	31,696	tS(GCU)	+
11,908	**tE(UUC)**	**+**	31,704	tS(GCU)	+
12,757	**tI(GAU)**	**-**	**31,711**	**tS(GCU)**	**+**
12,779	tI(GAU)	-	31,964	tS(UGA)	+
13,963	**tP(UGG)**	**-**	31,987	**tS(UGA)**	+
13,978	tP(UGG)	-	32,054	tL(UAG)	+
13,991	tP(UGG)	-	**32,070**	**tL(UAG)**	**+**
14,122	**tC(GCA)**	**-**	33,512	tM(CAU)2	+
14,128	tC(GCA)	-	33,539	**tM(CAU)2**	+

tRNAs. Unlike in plant mitochondria, where m6A modifications are abundant [64], early studies have indicated only low levels of methylation and pseudouridine on mitochondrial LSU rRNA in *S. cerevisiae* [65]. This likely also holds true for *C. albicans* rRNAs, and by extension also for coding transcripts. A single m6A in LSU rRNA was detected with high confidence, with two additional putative sites in the SSU rRNA and COB transcripts. These sites will require further experimental verification (such as reverse transcription with a modification-sensitive polymerase [66]). Once verified, the changes in these modification sites can be readily detected in future studies employing direct RNA sequencing on the ONT platform. Detecting other modification sites will however require a dedicated experimental approach, likely with a combination of immunoprecipitation with antibodies against a particular modified nucleotides along with ribodepletion and/or size selection of transcripts. The

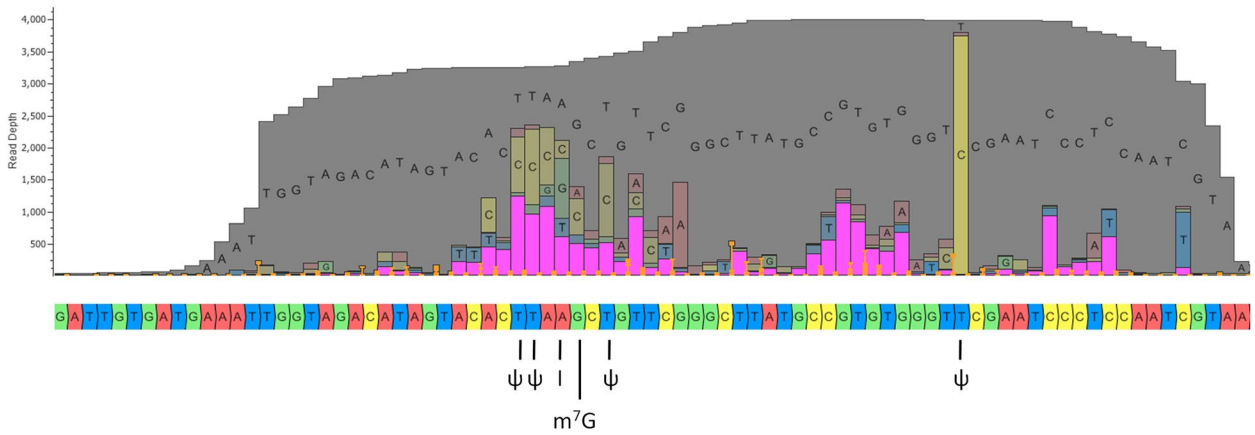

Fig. 6 Putative modifications detected in leucine tRNA (tL) by miscall analysis with ELIGOS. ψ – pseudouridine. I – inosine. m7G—N7-methylguanosine. Image has been generate with Golden Helix GenomeBrowser 3.1.0 [51]

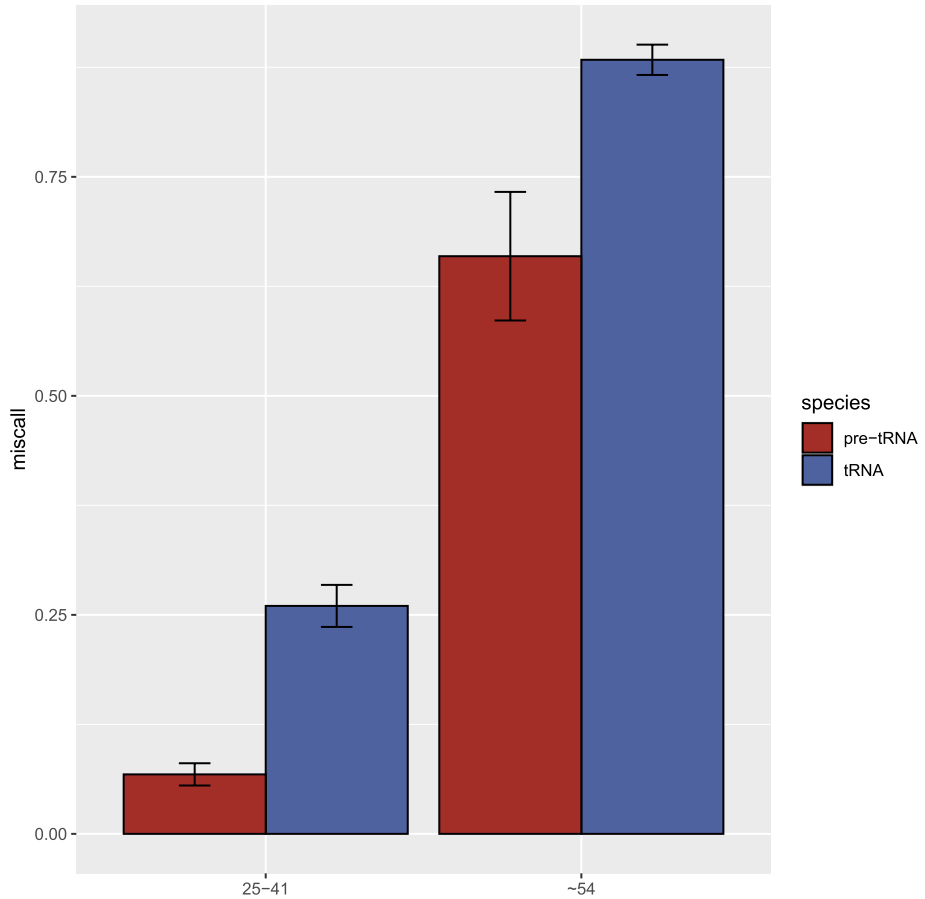

Fig. 7 Fraction of reads containing T-C miscalls for putative pseudouridine (ψ) sites in mature tRNAs and pre-tRNAs for position ~ 54 and positions 25–41 detected with ELIGOS. Error bars correspond to standard error. Presented data includes all putative ψ positions in 5 tRNAs present within TU03. This subset was selected due to relatively uniform pre-tRNA coverage. Comparison for all mitochondrial tRNAs yielded similar results and is presented in Supplementary Fig. 6

Table 6 Putative m5C (in blue) and pseudouridine (in red) sites detected by bisulfite sequencing. Each site was detected either by ELIGOS or NanoRSM analysis or both, either with three-step mapping or short read mapping

		three-step mapping		short read mapping	
position	gene	ELIGOS	NanoRSM	ELIGOS	NanoRSM
11558	tL(UAA)	+	+	+	+
11559	tL(UAA)	+			
11588	tL(UAA)	+		+	+
11671	tY(GUA)	+	+	+	+
11678	tY(GUA)	+	+	+	+
11756	tH(GUG)		+	+	+
11834	tT(UGU)		+	+	+
11908	tE(UUC)			+	+
14122	tC(GCA)			+	
12757	tI(GAU)				+
21207	tV(UAC)			+	+
21351	tP(UGG)			+	
31525	tQ(UUG)			+	+
31526	tQ(UUG)			+	+
31530	tQ(UUG)				+
31531	tQ(UUG)			+	+
31545	tQ(UUG)			+	+
31623	tD(GUC)	+		+	+
31705	tS(GCU)	+		+	+
31711	tS(GCU)			+	+

ribodepletion procedure for the *C. albicans* mitochondrial transcriptome also requires further optimization, as it should be applied to mitochondrial RNA of each investigated strain and sequenced separately alongside non-depleted RNA. Improved RNA004 chemistry recently introduced by ONT, which includes higher accuracy and throughput, along with modified nucleotide basecalling [67, 68] will likely also yield better quality and higher confidence results in future studies.

Conclusions

The results discussed above demonstrate, that direct RNA sequencing can reveal information about the mitochondrial transcriptome that would otherwise be unavailable via second-generation NGS. We believe that this preliminary study may serve as a proof of concept as well as a roadmap for analyzing direct RNA sequencing data from mitochondria of other yeast species and possibly mitochondrial transcriptomes from other groups of organisms with currently available bioinformatic tools. Due to their diversity in terms of mitochondrial genome organization and expression, along with a very broad range of evolutionary distances between individual species [69], yeasts are a particularly attractive model for mitochondrial comparative transcriptomic analysis, even more so with new types of data made available with ONT sequencing. Considering that nucleo-mitochondrial incompatibility has been found to be one of the mechanisms driving speciation in yeast [1, 70], a comparative analysis focusing on mitochondrial RNA modifications and nuclear proteins involved in this process appears particularly promising.

Abbreviations

IVT	*in vitro* Transcribed
TU	Transcription unit
RnI	Large subunit rRNA
Rns	Small subunit rRNA
ONT	Oxford Nanopore Technologies
MoP2	Master of Pores 2 data analysis pipeline
ELIGOS	Epitranscriptional/(Epigenomical) Landscape Inferring from Glitches of ONT Signals
RATTLE	Reference-free reconstruction and quantification of transcriptomes from long-read sequencing
LIQA	Long-read Isoform Quantification and Analysis
CHEUI	Methylation (CH3) Estimation Using Ionic current
CaPET127	*C. albicans* Ortholog of *S. cerevisiae* PET127 (PETite colonies). Systematic name: C1_11070W_A
CaDSS1	*C. albicans* Ortholog of *S. cerevisiae* DSS1 (Deletion of SUV3 Supressor). Systematic name: C2_08550C_A
mtEXO	Mitochondrial RNA degradosome
FTO	Fat mass and obesity-associated protein, an m6A RNA demethylase
m6A	N6-methyladenosine
m5C	5-Methylcytidine
m7G	N7-methylguanosine
Ψ	Pseudouridine
I	Inosine

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10791-4.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements
The authors would like to thank Michał Świrski and Katarzyna Iwańska for help in setting up the sequencing experiments, Dr Michał Marcinkowski for providing the FTO activity assay kit and Dr Agnieszka Maciejewska, Dr Maciej Kotliński and Jacek Olędzki for their support in conducting the MS-LC analysis.

Authors' contributions
J.P. performed the sequencing of native and synthetic C. albicans RNA, as well as performed the data analysis. K.K. performed the experimental part of bisulfite sequencing analysis. P.G. contributed to data interpretation and was a major contributor in manuscript revisions. All authors read and approved the final manuscript.

Funding
The work was supported by the National Science Centre, Poland, under research project no 2021/05/X/NZ2/01240.

Availability of data and materials
The datasets generated and analyzed during the current study, including raw ONT sequencing outputs, as well as all the scripts, commands, configuration and reference files used in the analysis are available in the Repository for Open Data (RepOD), at https://doi.org/https://doi.org/10.18150/EDN8ZG. Read data that support the findings of this study have also been deposited in the European Nucleotide Archive with the primary accession code PRJEB77681.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 July 2024 Accepted: 10 September 2024
Published online: 14 September 2024

References

1. Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. Cell. 2008. https://doi.org/10.1016/j.cell.2008.10.047.
2. Taylor JW, Berbee ML. Dating divergences in the fungal tree of life: review and new analyses. Mycologia. 2006;98:838–49. https://doi.org/10.3852/mycologia.98.6.838.
3. Dujon B. Yeast evolutionary genomics. Nat Rev Genet. 2010;11:512–24. https://doi.org/10.1038/nrg2811.
4. Jaehning JA. Mitochondrial transcription: is a pattern emerging? Mol Microbiol. 1993;8:1–4. https://doi.org/10.1111/j.1365-2958.1993.tb01197.x.
5. Rackham O, Mercer TR, Filipovska A. The human mitochondrial transcriptome and the rna-binding proteins that regulate its expression. Wiley Interdiscip Rev RNA. 2012;3:675–95. https://doi.org/10.1002/wrna.1128.
6. Verwilt J, Mestdagh P, Vandesompele J. Artifacts and biases of the reverse transcription reaction in RNA sequencing. RNA. 2023;29:889–97. https://doi.org/10.1261/rna.079623.123.
7. Ozsolak F, Milos PM. Single-molecule direct RNA sequencing without cDNA synthesis. Wiley Interdiscip Rev RNA. 2011. https://doi.org/10.1002/wrna.84.
8. Oikonomopoulos S, Bayega A, Fahiminiya S, Djambazian H, Berube P, Ragoussis J. Methodologies for transcript profiling using long-read technologies. Front Genet. 2020;11. https://doi.org/10.3389/fgene.2020.00606.
9. Liu H, Begik O, Lucas MC, Ramirez JM, Mason CE, Wiener D, et al. Accurate detection of m6A RNA modifications in native RNA sequences. Nat Commun. 2019. https://doi.org/10.1038/s41467-019-11713-9.
10. Tavakoli S, Nabizadehmashhadtoroghi M, Makhamreh A, Gamper H, Rezapour NK, Hou Y-M, et al. Detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct, long-read sequencing. bioRxiv. 2021.
11. Koster CC, Kleefeldt AA, van den Broek M, Luttik M, Daran JM, Daran-Lapujade P. Long-read direct RNA sequencing of the mitochondrial transcriptome of Saccharomyces cerevisiae reveals condition-dependent intron abundance. Yeast. 2023. https://doi.org/10.1002/yea.3893.
12. McShane E, Couvillion M, Ietswaart R, Prakash G, Smalec BM, Soto I, et al. A kinetic dichotomy between mitochondrial and nuclear gene expression drives OXPHOS biogenesis. bioRxiv Prepr Serv Biol. 2023;:2023.02.09.527880.
13. Łabędzka-Dmoch K, Kolondra A, Karpińska MA, Dębek S, Grochowska J, Grochowski M, et al. Pervasive transcription of the mitochondrial genome in Candida albicans is revealed in mutants lacking the mtEXO RNase complex. RNA Biol. 2021. https://doi.org/10.1080/15476286.2021.1943929.
14. Labedzka-Dmoch K, Razew M, Gapinska M, Piatkowski J, Kolondra A, Salmonowicz H, et al. The Pet127 protein is a mitochondrial 5′-to-3′ exoribonuclease from the PD-(D/E)XK superfamily involved in RNA maturation and intron degradation in yeasts. RNA. 2022. https://doi.org/10.1261/rna.079083.121.
15. Golik P. RNA processing and degradation mechanisms shaping the mitochondrial transcriptome of budding yeasts. IUBMB Life. 2024. https://doi.org/10.1002/iub.2779.
16. Wilson RB, Davis D, Mitchell AP. Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J Bacteriol. 1999. https://doi.org/10.1128/jb.181.6.1868-1874.1999.
17. Chiron S, Gaisne M, Guillou E, Belenguer P, Clark-Walker GD, Bonnefoy N. Studying mitochondria in an attractive model: Schizosaccharomyces pombe. Methods Mol Biol. 2007. https://doi.org/10.1007/978-1-59745-365-3_7.
18. Schmitt ME, Brown TA, Trumpower BL. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic Acids Res. 1990. https://doi.org/10.1093/nar/18.10.3091.
19. Zhu Y, Zhou G, Yu X, Xu Q, Wang K, Xie D, et al. LC-MS-MS quantitative analysis reveals the association between FTO and DNA methylation. PLoS ONE. 2017. https://doi.org/10.1371/journal.pone.0175849.
20. Mathur L, Jung S, Jang C, Lee G. Quantitative analysis of m6A RNA modification by LC-MS. STAR Protoc. 2021. https://doi.org/10.1016/j.xpro.2021.100724.
21. Shen L, Liang Z, Yu H. Dot Blot Analysis of N6-methyladenosine RNA modification levels. Bio-Protocol. 2017. https://doi.org/10.21769/bioprotoc.2095.
22. Amort T, Soulière MF, Wille A, Jia XY, Fiegl H, Wörle H, et al. Long non-coding RNAs as targets for cytosine methylation. RNA Biol. 2013. https://doi.org/10.4161/rna.24454.
23. Kim IV, Ross EJ, Dietrich S, Döring K, Sánchez Alvarado A, Kuhn CD. Efficient depletion of ribosomal RNA for RNA sequencing in planarians. BMC Genomics. 2019. https://doi.org/10.1186/s12864-019-6292-y.
24. Xu L, Berninger A, Lakin SM, O'Donnell V, Pierce JL, Pauszek SJ, et al. Direct RNA sequencing of foot-and-mouth disease virus genome using a flongle on MinION. Bio-Protoc. 2024;14:e5017.
25. Cozzuto L, Liu H, Pryszcz LP, Pulido TH, Delgado-Tejedor A, Ponomarenko J, et al. MasterOfPores: a workflow for the analysis of Oxford nanopore direct RNA sequencing datasets. Front Genet. 2020. https://doi.org/10.3389/fgene.2020.00211.
26. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford nanopore sequencing. Genome Biol. 2019. https://doi.org/10.1186/s13059-019-1727-y.

27. Anderson JB, Wickens C, Khan M, Cowen LE, Federspiel N, Jones T, et al. Infrequent genetic exchange and recombination in the mitochondrial genome of Candida albicans. J Bacteriol. 2001. https://doi.org/10.1128/JB.183.3.865-872.2001.
28. Wills JW, Troutman WB, Riggsby WS. Circular mitochondrial genome of Candida albicans contains a large inverted duplication. J Bacteriol. 1985. https://doi.org/10.1128/jb.164.1.7-13.1985.
29. Shaw JA, Troutman WB, Lasker BA, Mason MM, Riggsby WS. Characterization of the inverted duplication in the mitochondrial DNA of Candida albicans. J Bacteriol. 1989. https://doi.org/10.1128/jb.171.11.6353-6356.1989.
30. Kolondra A, Labedzka-Dmoch K, Wenda JM, Drzewicka K, Golik P. The transcriptome of Candida albicans mitochondria and the evolution of organellar transcription units in yeasts. BMC Genom. 2015. https://doi.org/10.1186/s12864-015-2078-z.
31. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018. https://doi.org/10.1093/bioinformatics/bty191.
32. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun. 2016. https://doi.org/10.1038/ncomms11307.
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009. https://doi.org/10.1093/bioinformatics/btp324.
34. Parker MT, Knop K, Barton GJ, Simpson GG. 2passtools: two-pass alignment using machine-learning-filtered splice junctions increases the accuracy of intron detection in long-read RNA sequencing. Genome Biol. 2021. https://doi.org/10.1186/s13059-021-02296-0.
35. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021. https://doi.org/10.1093/gigascience/giab008.
36. Lucas MC, Pryszcz LP, Medina R, Milenkovic I, Camacho N, Marchand V, et al. Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. Nat Biotechnol. 2024. https://doi.org/10.1038/s41587-023-01743-6.
37. Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics. 2021. https://doi.org/10.1093/bioinformatics/btaa692.
38. de la Rubia I, Srivastava A, Xue W, Indi JA, Carbonell-Sala S, Lagarde J, et al. RATTLE: reference-free reconstruction and quantification of transcriptomes from Nanopore sequencing. Genome Biol. 2022. https://doi.org/10.1186/s13059-022-02715-w.
39. Hu Y, Fang L, Chen X, Zhong JF, Li M, Wang K. LIQA: long-read isoform quantification and analysis. Genome Biol. 2021. https://doi.org/10.1186/s13059-021-02399-8.
40. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015. https://doi.org/10.1038/nmeth.3444.
41. Stoiber MH, Quick J, Egan R, Lee JE, Celniker SE, Neely R, et al. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. bioRxiv. 2016:094672. https://doi.org/10.1101/094672.
42. Leger A, Amaral PP, Pandolfini L, Capitanchik C, Capraro F, Miano V, et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. Nat Commun. 2021. https://doi.org/10.1038/s41467-021-27393-3.
43. Begik O, Lucas MC, Pryszcz LP, Ramirez JM, Medina R, Milenkovic I, et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. Nat Biotechnol. 2021. https://doi.org/10.1038/s41587-021-00915-6.
44. Jenjaroenpun P, Wongsurawat T, Wadley TD, Wassenaar TM, Liu J, Dai Q, et al. Decoding the epitranscriptional landscape from native RNA sequences. Nucleic Acids Res. 2021. https://doi.org/10.1093/nar/gkaa620.
45. Smith AM, Jain M, Mulroney L, Garalde DR, Akeson M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. PLoS ONE. 2019. https://doi.org/10.1371/journal.pone.0216709.
46. Xu L, Seki M. Recent advances in the detection of base modifications using the Nanopore sequencer. J Human Genet. 2020;65:25–33. https://doi.org/10.1038/s10038-019-0679-0.
47. Bozeman, MT: Golden Helix I. Golden Helix GenomeBrowse ® visualization tool (Version 3.1.0) [Software]. http://www.goldenhelix.com.
48. Mateos PA, Sethi AJ, Ravindran A, Srivastava A, Woodward K, Mahmud S, et al. Prediction of m6A and m5C at single-molecule resolution reveals a cooccurrence of RNA modifications across the transcriptome. bioRxiv. 2024;:2022.03.14.484124.
49. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. Nat Commun. 2019. https://doi.org/10.1038/s41467-019-08734-9.
50. Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, et al. Nanopore direct RNA sequencing maps the complexity of arabidopsis mRNA processing and m6A modification. Elife. 2020. https://doi.org/10.7554/eLife.49658.
51. Bozeman, MT: Golden Helix I. Golden Helix GenomeBrowse ® visualization tool (Version 3.1.0) [Software].
52. Thomas NK, Poodari VC, Jain M, Olsen HE, Akeson M, Abu-Shumays RL. Direct nanopore sequencing of individual full length tRNA strands. ACS Nano. 2021. https://doi.org/10.1021/acsnano.1c06488.
53. Ciampi MS, A-rena F, Cortese R, Daniel V. Biosynthesis of pseudouridine in the in vitro transcribed tRNATyr precursor. FEBS Lett. 1977. https://doi.org/10.1016/0014-5793(77)80196-8.
54. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. Nat Methods. 2019. https://doi.org/10.1038/s41592-019-0617-2.
55. Johnson Z, Xu X, Pacholec C, Xie H. Systematic evaluation of parameters in RNA bisulfite sequencing data generation and analysis. NAR Genomics Bioinforma. 2022. https://doi.org/10.1093/nargab/lqac045.
56. Slomovic S, Schuster G. Circularized RT-PCR (cRT-PCR): analysis of the 5′ends, 3′ends, and poly(A) tails of RNA. Methods Enzymol. 2013;530:227–51.
57. Cappannini A, Ray A, Purta E, Mukherjee S, Boccaletto P, Moafinejad SN, et al. MODOMICS: a database of RNA modifications and related information. 2023 update. Nucleic Acids Res. 2024;52:D239-44.
58. Lorenz C, Lünse CE, Mörl M. Trna modifications: Impact on structure and thermal adaptation. Biomolecules. 2017;7. https://doi.org/10.3390/biom7020035.
59. Song H, Zhang J, Liu B, Xu J, Cai B, Yang H, et al. Biological roles of RNA m5C modification and its implications in Cancer immunotherapy. Biomark Res. 2022;10. https://doi.org/10.1186/s40364-022-00362-8.
60. Zhao P, Xia L, Chen D, Xu W, Guo H, Xu Y, et al. METTL1 mediated tRNA m7G modification promotes leukaemogenesis of AML via tRNA regulated translational control. Exp Hematol Oncol. 2024. https://doi.org/10.1186/s40164-024-00477-8.
61. Rafels-Ybern À, Torres AG, Camacho N, Herencia-Ropero A, Frigolé HR, Wulff TF, et al. The expansion of inosine at the wobble position of tRNAs, and its role in the evolution of proteomes. Mol Biol Evol. 2019. https://doi.org/10.1093/molbev/msy245.
62. Gieg R, Eriani G. The tRNA identity landscape for aminoacylation and beyond. Nucleic Acids Res. 2023. https://doi.org/10.1093/nar/gkad007.
63. Suzuki T. The expanding world of tRNA modifications and their disease relevance. Nat Rev Mol Cell Biol. 2021;22:375–92. https://doi.org/10.1038/s41580-021-00342-0.
64. Murik O, Chandran SA, Nevo-Dinur K, Sultan LD, Best C, Stein Y, et al. Topologies of N6-adenosine methylation (m6A) in land plant mitochondria and their putative effects on organellar gene expression. Plant J. 2020. https://doi.org/10.1111/tpj.14589.
65. Klootwijk J, Klein I, Grivell LA. Minimal post-transcriptional modification of yeast mitochondrial ribosomal RNA. J Mol Biol. 1975. https://doi.org/10.1016/S0022-2836(75)80044-1.
66. Aschenbrenner J, Werner S, Marchand V, Adam M, Motorin Y, Helm M, et al. Engineering of a DNA Polymerase for Direct m6A sequencing. Angew Chemie - Int Ed. 2018. https://doi.org/10.1002/anie.201710209.
67. Liu-Wei W, van der Toorn W, Bohn P, Hölzer M, Smyth RP, von Kleist M. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. BMC Genomics. 2024;25:528.
68. White LK, Dobson K, Del Pozo S, Bilodeaux JM, Andersen SE, Baldwin A, et al. Comparative analysis of 43 distinct RNA modifications by nanopore tRNA sequencing. bioRxiv Prepr Serv Biol. 2024. https://doi.org/10.1101/2024.07.23.604651.

69. Sherman D, Durrens P, Beyne E, Nikolski M, Souciet JL. Génolevures: Comparative genomics and molecular evolution of hemiascomycetous yeasts. Nucleic Acids Res. 2004. https://doi.org/10.1093/nar/gkh091.
70. Jhuang H-Y, Lee H-Y, Leu J-Y. Mitochondrial-nuclear co-evolution leads to hybrid incompatibility through pentatricopeptide repeat proteins. EMBO Rep. 2017;18:87–101.

Publisher's Note