Mol Genet Genomics (2015) 290:1–9
DOI 10.1007/s00438-014-0905-6

REVIEW

The application of RNA-seq to the comprehensive analysis of plant mitochondrial transcriptomes

James D. Stone · Helena Storchova

Received: 9 May 2014 / Accepted: 21 August 2014 / Published online: 3 September 2014

Abstract We review current studies of plant mitochondrial transcriptomes performed by RNA-seq, highlighting methodological challenges unique to plant mitochondria. We propose ways to improve read mapping accuracy and sensitivity such as modifying a reference genome at RNA editing sites, using splicing- and ambiguity-competent aligners, and masking chloroplast- or nucleus-derived sequences. We also outline modified RNA-seq methods permitting more accurate detection and quantification of partially edited sites and the identification of transcription start sites on a genome-wide scale. The application of RNA-seq goes beyond genome-wide determination of transcript levels and RNA maturation events, and emerges as an elegant resource for the comprehensive identification of editing, splicing, and transcription start sites. Thus, improved RNA-seq methods customized for plant mitochondria hold tremendous potential for advancing our understanding of plant mitochondrial evolution and cytonuclear interactions in a broad array of plant species.

Communicated by S. Hohmann.

J. D. Stone · H. Storchova (✉)
Institute of Experimental Botany v.v.i, Academy of Sciences of the Czech Republic, Rozvojova 263, Prague 16502, Czech Republic
e-mail: storchova@ueb.cas.cz

J. D. Stone
Institute of Botany v.v.i, Academy of Sciences of the Czech Republic, Průhonice, Prague West 252 43, Czech Republic

Keywords RNA-seq · Plant mitochondria · Transcriptome · Editing

Introduction

RNA-seq has been shown to be an excellent tool for examining entire mitochondrial transcriptomes, capable of providing global information about transcript level and RNA maturation events, including editing. RNA-seq has only been applied to plant mitochondria in a few species, however. Nevertheless, its utility in plants is particularly promising, for characterizing the complexity of plant mitochondrial genomes and transcriptomes.

Mitochondrial genomes of higher plants are large (Sloan 2013), containing long intergenic regions and multiple promoters (Kuhn et al. 2005). They also undergo frequent intramolecular recombination, which may place coding sequences in the vicinity of previously unassociated regulatory elements (Kubo et al. 1999; Forner et al. 2005; Case and Willis 2008). Complex and frequent genomic rearrangements may be one reason why transcriptional control in plant mitochondria is rather relaxed, utilizing multiple promoters, sometimes with non-canonical motifs (Kühn et al. 2005; Zhang and Liu 2006). In addition to 24–41 protein coding genes, 3 rRNA, and 2–26 tRNA coding genes (Sloan et al. 2010a, 2012), various ORFs with unknown functions or putatively associated with cytoplasmic male sterility (CMS) have been shown to be transcribed (Handa et al. 1995; Terachi et al. 2001; Storchova et al. 2012; Okazaki et al. 2013).

In this review, we evaluate present applications of RNA-seq in plant mitochondrial transcriptome analyses. Based on these, as well as our own experience, we suggest methodological refinements which may further enhance the

application of RNA-seq techniques to plant mitochondrial transcriptomics.

Mitochondrial genome-wide transcription profiling

The application of high throughput sequencing to angiosperm mitochondrial transcriptomics has lagged behind the complete sequencing of plant mitochondrial genomes (reviewed in Sloan 2013). To date, researchers have used Illumina and SOLiD sequencing to examine mitochondrial transcriptomes in only a small handful of species (Table 1). The results confirm earlier conclusions obtained from genomic arrays in rice (Fujii et al. 2011). Notably, a high proportion (up to 49 %) of plant mitochondrial genomes is transcribed, albeit to a low extent, with isolated "islands" of higher transcriptional activity. Transcripts encoding respiratory chain proteins generally show higher abundances than transcripts encoding ribosomal proteins or cytochrome c biosynthesis enzymes. Mitochondrial transcript levels appear to be tissue- and organ-specific (Fang et al. 2012; Islam et al. 2013). However, the high variance among individual gene transcript levels, as well as the absence of biological replicates in these studies, makes it problematic to draw more specific conclusions. Accordingly, biologically replicated transcriptomic experiments are necessary to understand how mitochondrial gene expression is related to specific tissues, developmental stages and environmental conditions.

Genome-wide transcription profiling can also be used to reveal expression among uncharacterized mitochondrial ORFs. Four tobacco ORFs (*orf177, orf222, orf160, orf115*) were found to be transcribed and also associated with polysomes, suggesting they are translated (Grimes et al. 2014). Two of three expressed ORFs in *Brassica oleracea* (*orf266 and orf287*) were chimeric, containing portions of protein-coding genes (Grewe et al. 2014). Chimeric ORFs often encode CMS (Hu et al. 2014), but no CMS was detected in *B. oleracea*. It is therefore possible that the expression of chimeric ORFs in this plant is too low to cause CMS. In contrast, two novel transcripts, unique to male-sterile cytoplasms, may be associated with CMS in rubber tree (Shearman et al. 2014). Deep sequencing of plant mitochondrial transcriptomes makes it possible to identify genes or regions that are differentially expressed between male-sterile individuals and their male-fertile siblings sharing the same mitochondrial genome. If accompanied by a comparison of nuclear transcriptomes from the same individuals, this approach becomes a very convenient tool to better understand the cyto-nuclear causes and consequences of CMS in plants. To our knowledge, no such comprehensive study has been published.

Editing in plant mitochondria

RNA-seq also represents a convenient instrument for the global analysis of RNA editing, the post- or co-transcriptional mechanism that replaces select Cytidines with Uridines in mature mitochondrial and chloroplast transcripts (Gott and Emeson 2000; Knoop 2011). It has been primarily studied through the comparison of genomic and cDNA sequences corresponding to individual genes and obtained by Sanger sequencing. This approach is too laborious for the examination of all editing sites across organellar genomes and is not capable of identifying rarely edited sites ($\leq$10 % edited, Bentolila et al. 2013). RNA-seq offers solutions to both problems, allowing the simultaneous analysis of editing across all transcripts as well as quantifying editing extent (Picardi et al. 2010). Precise quantification of RNA editing is desirable as editing extent may vary in specific tissues or organs, be influenced by nucleo-cytoplasmic interactions (Hu et al. 2013), and contribute to CMS (Howad and Kempken 1997; Das et al. 2010).

Editing predominantly affects non-synonymous positions of protein-coding regions (Picardi et al. 2010; Grimes et al. 2014; Grewe et al. 2014), changing the resulting amino-acid sequences. tRNAs are edited to a lower extent-a single edit site (*trnC*) was found in cauliflower (Grewe et al. 2014), two tRNA edit sites (*trnC* and *trnF*) were detected in grapevine (Picardi et al. 2010), and five edits were identified in four tobacco tRNAs (Grimes et al. 2014). With the exception of *trnC*, tRNA edit sites are not conserved among plant species. No editing has been so far discovered in rRNA. Editing unlinked to any identifiable transcript is very rare-for example, only 7 of 540 editing sites in tobacco mitochondria fall into this category (Grimes et al. 2014).

The distribution of editing sites in plant mitochondrial genes is highly specific, controlled by complexes of editing proteins, confirming the functional importance of this process (Takenaka et al. 2014). Although somewhat conserved across angiosperms, there has been a trend of editing site loss during plant evolution. 1,782 Edits were found in the mitochondrial DNA of the lycophyte *Isoetes engelmannii* (Grewe et al. 2011), over 1,000 sites in *Cycas taitungensis* (Salmans et al. 2010), and more than 700 edits in a basal angiosperm *Liriodendron tulipifera* (Richardson et al. 2013). All of these are higher than editing site totals among the more-recently diverged eudicots and monocots, as summarized in Table 1. The mechanism of RNA editing site loss is not clear. One hypothesis posits retroprocessing, the reverse transcription and insertion of edited cDNA (Grewe et al. 2011). Another proposes C-to-T point mutations at the DNA level (Mower 2008; Sloan et al. 2010b). Deep sequencing of mitochondrial transcriptomes makes it possible to test the first hypothesis and to ask whether

Table 1 Recently published plant mitochondrial transcriptomic studies based on RNA-seq

Plant mt genome acc no	RNA	Extraction method (RNA size)	mRNA enrichment	Sequencing method	Number of reads generated	Read mapping method	Number of reads mapped	C-U editing definition	Editing sites identified (predicted)	Reference
Rubber tree AP014526	PolyA+	ND	Oligo dT priming	Illumina HiSeq2000 (parameters ND)	ND	Tophat 2.0.9, fusion search, Bowtie 1.0.0 (parameters ND)	ND	VarScan 2.3.4 (parameters ND)	224 (184)	Shearman et al. (2014)
Cauliflower KJ820683	Total	Tri-Reagent (Sigma)	ND	Illumina HiSeq, single 36 cycles	7.6 million	Bowtie 2.1.0	ND	>50 % of mapped reads	418	Grewe et al. (2014)
Tobacco NC006581	Total	Plant RNeasy (QIAgen) (>100 nt)	RiboZero (Epicentre)	Illumina HiSeq, PE 100 cycles	~164 million	DNASTAR SeqMan NGen (parameters ND)	4,539,709	DOC > 200 % edited < 100[a]	540	Grimes et al. (2014)
Ryegrass JX999996	Total	Plant RNeasy (QIAgen) (>100 nt)	ND	Illumina HiSeq, PE 100 cycles	ND	Bowtie (≤ 2 mismatches in first 25 bp), RSEM	ND	ND	ND	Islam et al. (2013)
Date Palm JN375330	Total	CTAB-based extraction	RiboMinus (Invitrogen)	SOLiD 50 nt	ND	BioScope 1.3 (parameters ND)	~7,00,000	ND	592	Fang et al. (2012)
Grapevine NC012119	PolyA+	ND	Oligo dT priming	Illumina 35 cycles	~205 million	PASS (90 % identity, ≥30 bp alignment)	9,39,554[b]	Fisher's exact test based on seq error estimation	401	Picardi et al. (2010)
	Total	ND	Oligo dT priming	SOLiD 35 nt	~328 million		5,207,827[b]			

[a] 100 % U-considered as SNP

[b] Only reads with unique target recovered

trancription level correlates with editing frequency, under the assumption that increased transcript levels may facilitate retro processing activity. Grewe et al. (2014) performed this test in tobacco with inconclusive results. However, a much broader taxonomic sampling will be required to fully evaluate this hypothesis, focusing particularly on mitochondria from generative apical meristems which are transferred to following generations.

Methods in plant mitochondrial transcriptomics

We summarize the current RNA-seq based plant mitochondrial transcriptomic studies in Table 1. We discuss the various methods used to prepare RNA, to enrich mRNA, to sequence cDNA libraries, to map the reads, and to analyze editing sites.

RNA preparation

Total cellular RNA or RNA isolated from purified mitochondria may be used for RNA-seq. The latter approach ensures that the vast majority of reads are derived from the mitochondrial genome and not from organellar sequences transferred to the nucleus (Michalovova et al. 2013) or the chloroplast genome. However, mitochondrial purification is accompanied by substantial losses and provides a low RNA yield (Leino et al. 2005), therefore total RNA was used as the starting material in existing studies. Both total and mitochondrial RNAs must be carefully treated with DNaseI to eliminate DNA contamination before sequencing. Most projects used popular adsorption column-based RNA extraction protocols. These methods fail to retain short RNAmolecules, however, and primarily transcripts larger than 100–200 nt are obtained. The resulting data should be interpreted as incomplete and lacking accurate information about short RNAs, including tRNAs and short non-coding RNAs, which may be derived from 3′and 5′ transcript ends (Ruwe and Schmitz-Linneweber 2012). This fact was not mentioned in any of the summarized plant transcriptomic studies.

Library preparation

Most stable mitochondrial mRNAs lack polyA tails (Gagliardi and Leaver 1999; Adamo et al. 2008), aspolyA tracts represent a signal for RNA degradation in plant mitochondria (Chang and Tong 2012; Hirayama et al. 2013). This role of polyA tails should be considered when selecting an appropriate cDNA synthesis method. OligodT priming, routinely used in nuclear transcriptomic studies, is not recommended for excluding rRNA neither from reverse transcription nor to enrich plant mitochondrial mRNA. Some reads derived from mitochondrial transcripts are always obtained, even with oligo dT primers, but they cannot be used for accurate transcript quantification. Such reads may be derived from polyadenylated RNA undergoing rapid degradation or they may result from random annealing of oligo dT to A-rich regions scattered across mitochondrial mRNAs (Stone and Storchova, unpublished). Nevertheless, methods other than oligodT priming of cDNA synthesis should be applied prior to reverse transcription to remove excess rRNA, comprising more than 95 % of total RNA in plants. Hybridization-based elimination of excessive rRNA (RiboZero, Invitrogen), for example, is a useful alternative to oligo dT priming. Early and recent mitochondrial transcriptomic studies (Picardi et al. 2010; Shearman et al. 2014) used oligodT priming, while another (Islam et al. 2013) did not explicitly mention the priming procedure. As oligodT priming primarily identifies polyA-tailed mRNA designated for degradation or the stretches near A-rich regions rather than accurately quantifying transcript abundance, the resulting transcriptomic data should be interpreted with caution.

Furthermore, in addition to enriching mRNA content, RNA must be carefully treated with DNAseI to eliminate DNA contamination before cDNA library preparation. Only Islam et al. (2013) and Grimes et al. (2014) explicitly mentioned this step, highlighting the need for standardized and transparent methods in future mitochondrial transcriptomic studies. Lastly, strand-specific libraries should be preferred over non-specific ones, because they distinguish sense- and anti-sense transcription, making the genomic source of transcripts more clear.

Sequencing

Picardi et al. (2010) compared Illumina and SOLiD sequencing methods, finding both suitable for plant mitochondrial transcriptomics. However, the reads were short (33–35 nt) and sequencing error rates relatively high, reducing read mapping accuracy and specificity. These shortcomings have largely been overcome by advances in sequencing technologies, particularly paired end sequencing and increased read lengths, up to 125 or 300 nt with the Illumina HiSeq or MiSeq platforms, respectively. As most plant mitochondrial genomes have a complex structure with numerous repeats far longer than read lengths, paired-end reads are particularly useful for increasing read mapping specificity and resolving repeats during *de novo* transcriptome assembly. This approach was followed in the recent transcriptomic study of tobacco (Grimes et al. 2014).

Mapping

Short reads are mapped to a reference sequence with the aid of an aligner. In the case of plant mitochondrial

transcriptomics, appropriate aligners should be capable of aligning reads across splice junctions and editing events. A genomic reference is required to positively identify mitochondrially-derived reads and was used in all of the studies reviewed in Table 1. Aligners typically use one of two strategies to compare reads with a reference: a burrows-wheeler transform (BWT) compressed index or seed hashing. BWT aligners are generally faster and use memory more efficiently, though this is not essential for mapping to a relatively small reference, such as plant mitochondria. Select hashing-based aligners offer greater sensitivity and compatibility with a reference containing ambiguity codes, eliminating bias for the reference base at an edited position. Due to the prevalence of RNA editing of mitochondrial transcripts, it is necessary to adjust aligner mismatch parameters or select an aligner that can interpret ambiguous bases to accurately map reads to their origins on a reference genome. Additional suggestions are made in the "Editing" section.

A reference mitochondrial genome may be adjusted to improve mapping quality. Genomic regions containing stretches of recently transferred chloroplast DNA require identification and possible masking to avoid mapping reads derived from chloroplast transcripts. The same approach may be applied to sequences of nuclear origin if they are positively identified. Labeling non-uniquely mapped reads, as the aligner allows, highlights both repeated regions, as well as the unique features that strongly influence the allocation of non-uniquely mapping reads. This information is vital for accurate transcript quantification within an RNA fragments' length of repeated sequences. Figure 1 compares mapping results obtained with and without allowing multiple read matches.

Transcript level estimation

The number of transcription unit (TU)-specific reads normalized by total read count and TU length cannot be interpreted as a direct measure of gene expression. Unlike the constant number of nuclear genes present in a single cell, growing evidence indicates that not only the number of mitochondria per cell, but also gene copy number per mitochondrion are highly variable (Woloczynska et al. 2006; Preuten et al. 2010). Thus, one should consider both read coverage and gene copy number to obtain realistic estimates of mitochondrial gene expression. The relative copy number of mitochondrial genomic regions of interest should be estimated from total DNA of the same tissue specimen which was used to prepare total RNA, by means of quantitative PCR or from high throughput sequencing data sets. The copy number ratio should be then used to correct the normalized coverage of the specific TU to achieve realistic transcript level values. Such a comprehensive analysis has not been performed in any cited paper, although the necessity of considering mitochondrial gene copy number in estimating gene expression levels was briefly discussed by Fang et al. (2012).

Editing

Applying RNA-seq to the detection and quantification of RNA editing requires that reads derived from edited transcripts are accurately mapped to a reference genome they no longer match. Outside of a few coding sequences (CDS), editing density is sufficiently low that multiple RNA editing events rarely occur within the same read. Therefore, outside of CDS, an aligner can adequately map edited

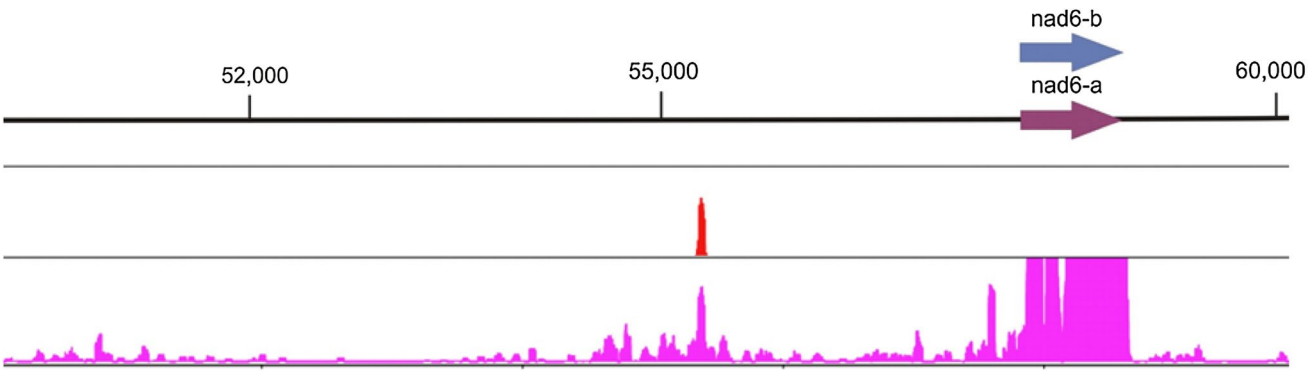

Fig. 1 SOLiD reads derived from cDNA and mapped against the mitochondrial genome of *Silene vulgaris* JQ771300 using CLC Genomics Server v.4.9 (CLC Bio). The first box shows the transcription profile obtained from uniquely matching reads, the second one displays the results of mapping reads with multiple matches. The *nad6* gene exists in two copies in the mitochondrial genome, its 5′ flanking region is also repeated and transcribed at low level. There is not sufficient information to determine whether one or the two *nad6* copies and their adjacent regions are expressed. A single prominent peak close to the 55,000 position corresponds to a short unique sequence motif, specific to only one of the two *nad6* copies

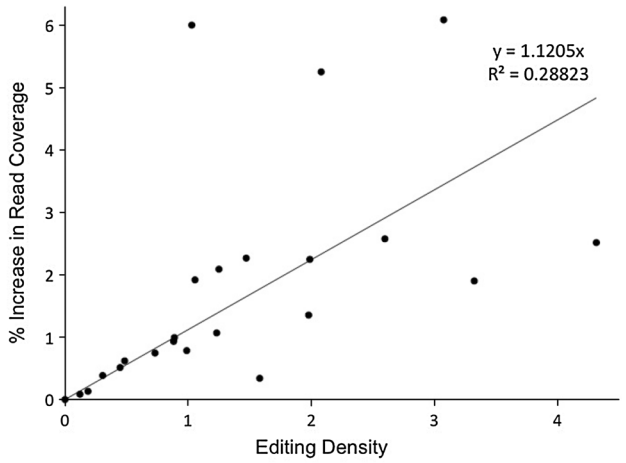

Fig. 2 Increase in read mapping coverage vs. editing density for a modified reference genome where all major editing sites were changed to their edited (T) form. The data points correspond to coding sequences for all 24 single-copy protein coding genes in the *Silene vulgaris* mitochondrion. *nad6-a*, *nad6-b*, *mttB-a*, and *mttB-b* were excluded as their homology precluded accurate read mapping to either copy. 50 nt SOLiD reads were mapped separately to unedited and edited reference genomes using CLC Genomics Workbench v.6.5 (CLC Bio) using stringent parameters. The percent change in coverage was calculated as the difference in coverage between the edited and unedited mappings, divided by the unedited coverage, and multiplied by 100 %. Editing density was calculated as the number of highly edited positions per CDS length, multiplied by 100 %. Outlier data may be explained by non-uniformly distributed editing sites or the presence of partially edited sites, reducing increases in coverage. Similar results were obtained using additional aligners (data not shown)

reads to an unmodified genomic reference sequence, with a slight relaxation of mismatch settings. As editing density increases, or where editing sites are clustered, reads are unlikely to map to the genomic reference without allowing for more mismatches, an associated reduction in specificity, and incorporation of sequencing errors (Lee et al. 2013). To accurately quantify RNA editing, especially within highly edited CDS, we have found it necessary to modify the reference genome at putative editing sites (Fig. 2) and to supply this information to a variant-capable aligner.

An initial run of the mapping procedure with relaxed mismatch penalties, followed by variant calling, may be used to identify well-supported editing sites, i.e. high coverage and high editing extent. The predictive RNA editing tool, PREP-Mt (Mower 2009) may also be used to identify candidate RNA editing sites within coding sequences. These putative edited sites may then be incorporated into the reference genome as an ambiguous base (Y or R, depending on transcribed strand), or as a supplemental file of known variants, allowing both unedited and edited reads to align during a final mapping procedure. A modified reference and aligner capable of interpreting ambiguous bases are necessary to quantify partial editing without bias (Ruwe et al. 2013). Similar methods have been proposed to enhance detection of allelic variants (Satya et al. 2012; Stevenson et al. 2013) and methylated bases through bisulfate sequencing (Lim et al. 2012), though these generally lack support for splicing and for correctly mapping edits that revert a Uridine to a Cytidine. Bowtie2 (Langmead and Salzberg 2012), a widely used BWT aligner, scores all bases aligning to an ambiguous reference position as a unique category of mismatch and permits the user to set the mismatch penalty for alignment with ambiguous bases to zero. This removes bias towards the genomic base (Satya et al. 2012) at edited positions and facilitates mapping of both edited and unedited reads, albeit with the trade-off of allowing sequencing errors to align to the modified base. Bowtie2 also benefits from compatibility with the splice junction aligner, TopHat2 (Kim et al. 2013), enabling accurate mapping around both splice junctions and editing sites.

The GSNAP aligner (Wu and Nacu 2010) also offers support for splicing and allows the user to specify multiple bases at edited positions in the reference sequence. Furthermore, GSNAP has been designed with bisulfate sequencing in mind and can distinguish C to T and G to A alignments from other mismatches, without the assistance of a modified genome. The STAR aligner (Dobin et al. 2013), based on uncompressed suffix-array seed-extension, also offers native support for mapping reads across splice junctions. STAR lacks ambiguity code support, however, requiring relaxed mismatch penalties to ensure that edited bases successfully align to a genomic reference. MOSAIK (Lee et al. 2014) and Novoalign (Novocraft Technologies) are hashed seed-extension based aligners with full support for IUPAC ambiguity codes, aligning both edited and unedited bases to a modified reference without incorporating sequencing errors at modified bases like Bowtie2 and STAR. These aligners lack support for splice junctions, however, decreasing mapping coverage and accuracy near splicing events. Selecting the ideal aligner depends on the project's aims. If the goal is limited to accurate alignment and RNA editing quantification among protein coding genes, the difficulties of mapping around splice junctions can be alleviated by using CDS reference sequences. In this case, modifying the CDS reference at putative editing sites, using an aligner competent to interpret ambiguity codes (GSNAP, Novoalign, MOSAIK), and applying stringent parameters should give the most accurate results. If the goals include genome-wide transcript and editing quantification, a splice junction aligner (Tophat2 using Bowtie2, GSNAP, STAR) and modified reference are advised.

More sensitive approaches are necessary to identify and accurately quantify editing at sites that are partially edited and/or outside of CDS where coverage is low and predictive tools based on amino acid sequence conservation (Mower 2009) cannot be used. Under these

circumstances, base recalibration of aligned reads (Van der Auwera et al. 2013, updated on http://www.broadinstitute.org/gatk/guide/best-practices) can be effective at reducing the number of false positives due to sequencing error and inaccurate base call qualities. The Base Recalibrator tool of GATK (DePristo et al. 2011) uses a user-supplied list of known editing sites and an alignment, both of which may be generated from an initial mapping procedure, to identify systematic biases in base quality scores that contribute to mismatches that do not fit the standard RNA editing pattern. The resulting recalibration data is used to revise base qualities in the alignment prior to a final round of variant calling. Alternatively, Bentolila et al. (2013) developed a method to accurately quantify editing in organelles by Illumina sequencing, while accounting for empirically estimated mismatch rates. As deep, uniform coverage is required for precise measurement of editing extent, RT-PCR amplicons derived from genic regions were mixed in an equimolar ratio before sequencing.

Future perspectives

Another promising application named dRNA-seq (differential RNA sequencing), has been used so far only in chloroplasts. This approach enables the determination of all transcription start sites across the organellar genome (Zhelyazkova et al. 2012). Before cDNA library preparation and high throughput sequencing, RNA is treated with terminator exonuclease (TEX). This enzyme degrades only the transcripts with a 5′monophosphate which are products of post-transcriptional processing. Primary transcripts terminated with 5′-triphosphates remain intact. Transcription starts are then marked by a sharp vertical boundary of mapped reads visible in a global transcription landscape after TEX treatment. This novel approach has the potential to replace the cumbersome single-gene targeted method of RNA-circularization (Kühn and Binder 2002; Zhang and Liu 2006; Forner et al. 2007; Müller and Storchova 2013).

Mammalian mitochondrial genomes are very different from their angiosperm counterparts. They are small, compact, with a single promoter region responsible for a polycistronic transcript which undergoes complex post-transcriptional processing (Anderson 1981). Despite these differences, a recent comprehensive analysis of the human mitochondrial transcriptome (Mercer et al. 2011) may inspire similar approaches in plant mitochondrial studies. The authors used pure mitochondria and mitoplasts (mitochondria without outer membranes) for RNA extraction. This method allowed the identification of nuclear-encoded non-coding RNA which had been transferred to mitochondria. The processing of polycistronic precursors was examined using parallel analysis of RNA ends (PARE - German et al. 2008), which identifies the 5′ends of cleaved transcripts based on their exposed 5′-monophosphate. Deep analysis of in vivo DNaseI-cleaved DNA fragments made it possible to recognize genomic stretches protected by proteins and led to the construction of a map of putative protein-DNA interaction sites in the human mitochondrial genome.

Thus RNA-seq, creatively combined with other methods developed in chloroplast or in animal mitochondrial studies, has enormous potential to deepen our understanding of complicated processes responsible for transcriptional control in plant mitochondria.

Acknowledgments We are grateful to Daniel B Sloan for reading and commenting on this manuscript. This work originated in the framework of the project "Integration of the experimental and population biology using new methods of interdisciplinary issues-the way to excellence with young scientists," Reg.No.: CZ.1.07/2.3.00/30.0048, funded by the European Social Fund (ESF) and the state budget of Czech Republic through the Operational Programme Education for Competitiveness (OPEC). It was further supported by the grant of the Grant Agency of the Czech Republic P506/12/1359 to H.S.

Conflict of interest The authors declare that they have no conflict of interest.

References

Adamo A, Pinney JW, Kunova A, Westhead DR, Meyer P (2008) Heat stress enhances the accumulation of polyadenylated mitochondrial transcripts in Arabidopsis thaliana. PLoS One 3:e2889. doi:10.1371/journal.pone.0002889

Anderson L (1981) Identification of mitochondrial proteins and some of their precursors in two-dimension a electrophoretic maps of human cells. Proc Natl Acad Sci USA 78:2407–2411. doi:10.1073/pnas.78.4.2407

Bentolila S, Oh J, Hanson MR, Bukowski R (2013) Comprehensive high-resolution Analysis of the role of an Arabidopsis gene family in RNA editing. PLoS Genet 9:e1003584. doi:10.1371/journal.pgen.1003584

Case AL, Willis JH (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. Evolution 62:1026–1039. doi:10.1111/j.1558-5646.2008.00360.x

Chang JH, Tong L (2012) Mitochondrial poly(A) polymerase and polyadenylation. Biochim Biophys Acta 1819:992–997. doi:10.1016/j.bbagrm.2011.10.012

Das S, Sen S, Chakraborty A, Chakraborti P, Maiti MK, Basu A, Basu D, Sen SK (2010) An unedited 1.1 kb mitochondrial orfB gene transcript in the Wild Abortive Cytoplasmic Male Sterility (WA-CMS) system of Oryza sativa L. subsp indica. BMC Plant Biol 10:39. doi:10.1186/1471-2229-10-39

DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. doi:10.1038/ng.806

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. doi:10.1093/bioinformatics/bts635

Fang Y, Wu H, Zhang T, Yang M, Yin Y, Pan L, Yu X, Zhang X, Hu S, Al-Mssallem IS, Yu J (2012) A complete sequence and transcriptomic analyses of date palm (Phoenix dactylifera L.) mitochondrial genome. PLoS One 7:e37164. doi:10.1371/journal.pone.0037164

Forner J, Weber B, Wietholter C, Meyer RC, Binder S (2005) Distant sequences determine 5′ end formation of *cox3* transcripts in *Arabidopsis thaliana* ecotype C24. Nucleic Acids Res 33:4673–4682. doi:10.1093/nar/gki774

Forner J, Weber B, Thuss S, Wildum S, Binder S (2007) Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5′and 3′ end formation. Nucleic Acids Res 35:3676–3692. doi:10.1093/nar/gkm270

Fujii S, Toda T, Kikuchi S, Suzuki R, Yokoyama K, Tsuchida H, Yano K, Toriyama K (2011) Transcriptome map of plant mitochondria reveals islands of unexpected transcribed regions. BMC Genom 12:279. doi:10.1186/1471-2164-12-279

Gagliardi D, Leaver JCJ (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. EMBO J 18:3757–3766. doi:10.1093/emboj/18.13.3757

German MA, Pillay M, Jeong DH, Hetawal A, Luo SJ, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R et al (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNAends. Nat Biotechnol 26:941–946. doi:10.1038/nbt1417

Gott JM, Emeson RB (2000) Functions and mechanisms of RNA editing. Annu Rev Genet 34: 499–U34. doi:10.1146/annurev.genet.34.1.499

Grewe F, Herres S, Viehöver P, Polsakiewicz M, Weisshaar B, Knoop V (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. Nucl Acids Res 39:2890–2902. doi:10.1093/nar/gkq1227

Grewe F, Edger PP, keren I, Sultan L, Pires JC, Ostersetzer-Biran O, Mower JP (2014) Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. Mitochondrion. doi:10.1016/j.mito.2014.05.008

Grimes BT, Sisay AK, Carroll HD, Cahoon AB (2014) Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding region. BMC Genom 15:31. doi:10.1186/1471-2164-15-31

Handa H, Gualberto JM, Grienenberg JM (1995) Characterization of the mitochondrial orfB gene and its derivative, *orf224*, a chimeric open reading frame specific to one mitochondrial genome of the Polima male-sterile cytoplasm in rapeseed (Brassica napus L.). Curr Genet 28:546–552. doi:10.1007/BF00518167

Hirayama T, Matsuura T, Ushiyama S, Narusaka M, Kurihara Y, Yasuda M, Ohtani M, seki M, demura T, Nakashita H, Narusaka Y, Hayashi S (2013) A poly(A)-specific ribonuclease directly regulates the poly(A) status of mitochondrial mRNA in Arabidopsis. Nat Commun 4:2247. doi:10.1038/ncomms3247

Howad W, Kempken F (1997) Cell type-specific loss of at p6 RNA editing in cytoplasmic male sterile Sorghum bicolor. Proc Natl Acad Sci USA 94:11090–11095. doi:10.1073/pnas.94.20.11090

Hu JH, Yi R, Zhang HY, Ding Y (2013) Nucleo-cytoplasmic interactions affect RNA editing of cox2, atp6 and atp9 in alloplasmic male-sterile rice (Oryza sativa L.) lines. Mitochondrion 13:87–95. doi:10.1016/j.mito.2013.01.011

Hu J, Huang W, Huang Q, Qin X, Yu C, Wang L, Li S, Zhu R, Zhu Y (2014) Mitochondria and cytoplasmic male sterility in plants. Mitochondrion. doi:10.1016/j.mito.2014.02.008

Islam MS, Studer B, Byrne SL, Farrell JD, Panitz F, Bendixen C, Moller IA, Asp T (2013) The genome and transcriptome of perennial ryegrass mitochondria. BMC Genom 14:202. doi:10.1186/1471-2164-14-202

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. doi:10.1186/gb-2013-14-4-r36

Knoop V (2011) When you can't trust the DNA: rNA editing changes transcript sequences. Cell Mol Life Sci 68:567–586. doi:10.1007/s00018-010-0538-9

Kubo T, Nishizawa S, Mikami T (1999) Alterations in organization and transcription of the mitochondrial genome of cytoplasmic male sterile sugar beet (*Beta vulgaris L.*). Mol Gen Genet 262:283–290. doi:10.1007/s004380051085

Kühn J, Binder S (2002) RT-PCR analysis of 5′ to 3′ end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. Nucleic Acids Res 30:439–446. doi:10.1093/nar/30.2.439

Kühn K, Weihe A, Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. Nucleic Acids Res 33:337–346. doi:10.1093/nar/gki179

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi:10.1038/nmeth.1923

Lee J-H, Ang JK, Xiao X (2013) Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. RNA 19:725–732. doi:10.1261/rna.037903.112

Lee W-P, Stromberg MP, Ward A, Steward C, Garrison EP, Marth GT (2014) MOSAIK: a Hash-Based Algorithm for Accurate Next-Generation Sequencing Short-Read Mapping. PLoS ONE 9(3):e90581. doi:10.1371/journal.pone.0090581

Leino M, Landgren M, Glimelius K (2005) Alloplasmic effects on mitochondrial transcriptional activity and RNA turnover result in accumulated transcripts of *Arabidopsis* orfs in cytoplasmic male-sterile *Brassica napus*. Plant J 42:469–480. doi:10.1111/j.1365-313X.2005.02389.x

Lim J-Q, Tennakoon C, Li G, Wong E, Ruan Y, Wei C-L, Sung W-K (2012) BatMeth: improved mapper for bisulfite sequencing reads on DNA methylation. Genome Biol 13:R82. doi:10.1186/gb-2012-13-10-r82

Mercer TM, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood AMJ, Haugen E, Bracken CP, Rackham O, Stamatoyannopoulos JA, Filipovska A, Mattick JS (2011) The human mitochondrial transcriptome. Cell 146:645–658. doi:10.1016/j.cell2011.06.051

Michalovova M, Vyskot B, Kejnovsky E (2013) Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity 111:314–320. doi:10.1038/hdy.2013.51

Mower JP (2008) Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. Mol Biol Evol 25:52–61. doi:10.1093/molbev/msm226

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253–W259. doi:10.1093/nar/gkp337

Müller K, Storchova H (2013) Transcription of atp1 is influenced by both genomic configuration and nuclear background in the highly rearranged mitochondrial genomes of Silene vulgaris. Plant Mol Biol 81:495–505. doi:10.1007/s11103-013-0018-3

Okazaki M, Kazama T, Murata H, Motomura K, Toriyama K (2013) Whole mitochondrial genome sequencing and transcriptional analysis to uncover an rt102-type cytoplasmic male sterility-associated candidate gene derived from *Oryza rufipogon*. Plant Cell Physiol 54:1560–1568. doi:10.1093/pcp/pct102

Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G (2010) Large-scale detection and analysis of RNA editing in

grape mtDNA by RNA deep-sequencing. Nucleic Acids Res 38:4755–4767. doi:10.1093/nar/gkq202

Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Boerner T (2010) Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. Plant J64:948–959. doi:10.1111/j.1365-313X.2010.04389.x

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD (2013) The "fossilized" mitochondrial genome of Liriodendron tulipifera: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biol 11:29. doi:10.1186/1741-7007-11-29

Ruwe H, Schmitz-Linneweber C (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? Nucleic Acids Res 40:3106–3116. doi:10.1093/nar/gkr1138

Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB (2013) Arabidopsis chloroplast quantitative editotype. FEBS Lett 587:1429–1433. doi:10.1016/j.febslet.2013.03.022

Salmans ML, Chaw SM, Lin CP, Shih ACC, Wu YW, Mulligan RM (2010) Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. Curr Genet 56:439–446. doi:10.1007/s00294-010-0312-4

Satya RV, Zavaljevski N, Reifman J (2012) A new strategy to reduce allelic bias in RNA-Seq readmapping. Nucleic Acids Res 40:e127. doi:10.1093/nar/gks425

Shearman JR, Sangsrakru D, Ruang-Areerate P, Sonthirod C, Uthaipaisanwong P, Yoocha T, Poopear S, Theerawattanasuk K, Tragoonrung S, Tangphatsornruang S (2014) Assembly and analysis of a male sterile rubber tree mitochondrial genome reveals DNA rearrangement events and a novel transcript. BMC Plant Biol 14:45. doi:10.1186/1471-2229-14-45

Sloan DB (2013) One ring to rule them all? Genome sequencing provides new insights into the 'master circle'model of plant mitochondrial DNA structure. New Phytol 200:978–985. doi:10.1111/nph.12395

Sloan DB, Alverson AJ, Štorchová H, Palmer JD, Taylor DR (2010a) Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evol Biol 10:274. doi:10.1186/1471-2148-10-274

Sloan DB, MacQueen AH, Alverson AJ, almer JD, Taylor DR (2010b) Extensive loss of rna editing sites in rapidly evolving silene mitochondrial genomes: selection vs. retroprocessing as the driving force. Genetics 185: 1369-U358. doi:10.1534/genetics.110.118000

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10:e1001241. doi:10.1371/journal.pbio.1001241

Stevenson KR, Coolon JD, Wittkopp PJ (2013) Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. BMC Genom 14:536. doi:10.1186/1471-2164-14-536

Storchova H, Müller K, Lau S, Olson MS (2012) Mosaic origin of a complex chimeric mitochondrial gene in *Silene vulgaris*. PLoS One 7:e30401. doi: 10.1371/journal.pone.0030401

Takenaka M, Verbitskiy D, Zehrmann A, Hartel B, Bayer-Csaszar E, Glass F, Brennicke A (2014) RNA editing in plant mitochondria—Connecting RNA target sequences and acting proteins. Mitochondrion. doi:10.1016/j.mito.2014.04.005

Terachi T, Yamaguchi K, Yamagishi H (2001) Sequence analysis on the mitochondrial orfB locus in normal and Ogura male-sterile cytoplasms from wild and cultivated radishes. Curr Genet 40:276–281. doi:10.1007/s00294-001-0256-9

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013) From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinform 43:11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43

Woloczynska M, Kmiec B, Mackiewicz P, Janska H (2006) Copy number of bean mitochondrial genes estimated by real-time PCR does not correlate with the number of gene loci and transcript levels. Plant Mol Biol 61:1–12. doi:10.1007/s11103-005-5773-3

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881. doi:10.1093/bioinformatics/btq057

Zhang QY, Liu YG (2006) Rice mitochondrial genes are transcribed by multiple promoters that are highly diverged. J Integr Plant Biol 48:1473–1477. doi:10.1111/j.1672-9072.2006.00384.x

Zhelyazkova P, Sharma CM, Fortsner KU, Liere K, Vogel J, Boerner T (2012) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. Plant Cell 24:123–136. doi:10.1105/tpc.111.089441