Molecular Phylogenetics and Evolution 98 (2016) 57–62

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Recovering complete mitochondrial genome sequences from RNA-Seq: A case study of *Polytomella* non-photosynthetic green algae ☆

Yao Tian, David Roy Smith *

Department of Biology, University of Western Ontario, London, ON N6A 5B7, Canada

ARTICLE INFO

Article history:
Received 29 May 2015
Accepted 28 January 2016
Available online 6 February 2016

Keywords:
Marine Microbial Eukaryotic Transcriptome Sequencing Project
Mitochondrial genome
Next-generation sequencing
Polytomella
Mitochondrial transcriptome
Telomere

ABSTRACT

Thousands of mitochondrial genomes have been sequenced, but there are comparatively few available mitochondrial transcriptomes. This might soon be changing. High-throughput RNA sequencing (RNA-Seq) techniques have made it fast and cheap to generate massive amounts of mitochondrial transcriptomic data. Here, we explore the utility of RNA-Seq for assembling mitochondrial genomes and studying their expression patterns. Specifically, we investigate the mitochondrial transcriptomes from *Polytomella* non-photosynthetic green algae, which have among the smallest, most reduced mitochondrial genomes from the Archaeplastida as well as fragmented rRNA-coding regions, palindromic genes, and linear chromosomes with telomeres. Isolation of whole genomic RNA from the four known *Polytomella* species followed by Illumina paired-end sequencing generated enough mitochondrial-derived reads to easily recover almost-entire mitochondrial genome sequences. Read-mapping and coverage statistics also gave insights into *Polytomella* mitochondrial transcriptional architecture, revealing polycistronic transcripts and the expression of telomeres and palindromic genes. Ultimately, RNA-Seq is a promising, cost-effective technique for studying mitochondrial genetics, but it does have drawbacks, which are discussed. One of its greatest potentials, as shown here, is that it can be used to generate near-complete mitochondrial genome sequences, which could be particularly useful in situations where there is a lack of available mtDNA data.

1. Introduction

For more than three decades scientists have been sequencing mitochondrial genomes. So much so that mitochondrial DNAs (mtDNAs) are now among the most sequenced and publicized type of chromosome (Smith and Keeling, 2015). Other aspects of mtDNAs have also been intensely investigated, including their conformations and modes of gene expression (Fan et al., 2003). Unfortunately, however, research on features of mitochondrial chromosomes apart from the primary DNA sequence have not kept pace with genome sequencing (Smith, 2016). But this could soon be changing.

Massively parallel RNA-sequencing (RNA-Seq) technologies provide a cheap, efficient means for exploring gene expression (Metzker, 2010). Although widely used to study eukaryotic nuclear transcription, RNA-Seq is an excellent but untapped resource for investigating mitochondrial genetics (Smith, 2013; Hodgkinson et al., 2014). The high copy number per cell and elevated expression levels of mtDNA mean that mitochondrial-derived transcripts can represent a large fraction of the reads generated from eukaryotic RNA-Seq experiments (Raz et al., 2011; Smith, 2013). Intergenic regions of mitochondrial genomes can also be transcriptionally active (Barbrook et al., 2010; Rackham et al., 2011), permitting the recovery of both coding and noncoding sequences from RNA-Seq results.

Given the popularity of RNA-Seq within life science research, public online genetic databanks, such as the National Centre for Biotechnology Information (NCBI), are accumulating huge amounts of raw RNA-Seq data from diverse eukaryotes. As of April 1, 2015, NCBI's Sequence Read Archive contained 3.4 quadrillion bases of high-throughput sequencing data, much of which are RNA-Seq. In many cases, the researchers that generated these data ignored the mitochondrial sequences (Smith, 2012, 2013), meaning that NCBI harbors billions of mitochondrial-derived reads from hundreds of different eukaryotes just waiting to be assembled and analyzed. Moreover, the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) made publicly available the transcriptomes from hundreds of diverse protists (Keeling et al., 2014). The raw Illumina sequencing data from these transcriptomes, which are in NCBI, represent an exceptional opportunity

☆ This paper was edited by the Associate Editor Elizabeth Zimmer.

* Corresponding author.
E-mail address: dsmit242@uwo.ca (D.R. Smith).

http://dx.doi.org/10.1016/j.ympev.2016.01.017
1055-7903/

for studying organelle transcription from some of the most poorly studied but ecologically important organisms on Earth.

One group elected for transcriptome sequencing by the MMETSP was *Polytomella*—a monophyletic green algal genus of free-living, freshwater, non-photosynthetic unicells, closely related to *Chlamydomonas reinhardtii* (Nakada et al., 2008; Smith et al., 2013). There are presently four known *Polytomella* lineages, represented by *Polytomella piriformis*, *P. parva*, *P. capuana*, and *P. magna*. The mtDNAs of these four species have been studied in great detail (Fig. 1A) (Fan and Lee, 2002; Smith and Lee, 2008; Smith et al., 2010, 2013). They are small (~13–28 kb), made up of 1–2 linear chromosomes with invertedly repeated telomeres, have matching or very similar gene orders, which are organized into opposing transcriptional units, and contain identical gene contents, encoding 7 standard mitochondrial proteins, 1 tRNA, and 2 rRNAs. Moreover, in *P. magna*, the deepest-branching member of the genus, every gene in the mitochondrial genome is duplicated and in an inverted orientation relative to its partner (Fig. 1A) (Smith et al., 2013). Together, these unusual genomic features make *Polytomella* an excellent candidate for studying organelle gene expression.

Here, we use Illumina RNA-Seq to examine the mitochondrial transcriptomes of the four known *Polytomella* lineages. Our goal is to unravel the roles, if any, of the *Polytomella* mitochondrial telomeres and palindromic elements in gene expression, and to characterize aspects of posttranscriptional processing. In a broader sense, we want to explore the effectiveness of RNA-Seq for recovering entire mitochondrial genomes. In the end, we find many benefits as well as some limitations to using RNA-Seq for organelle

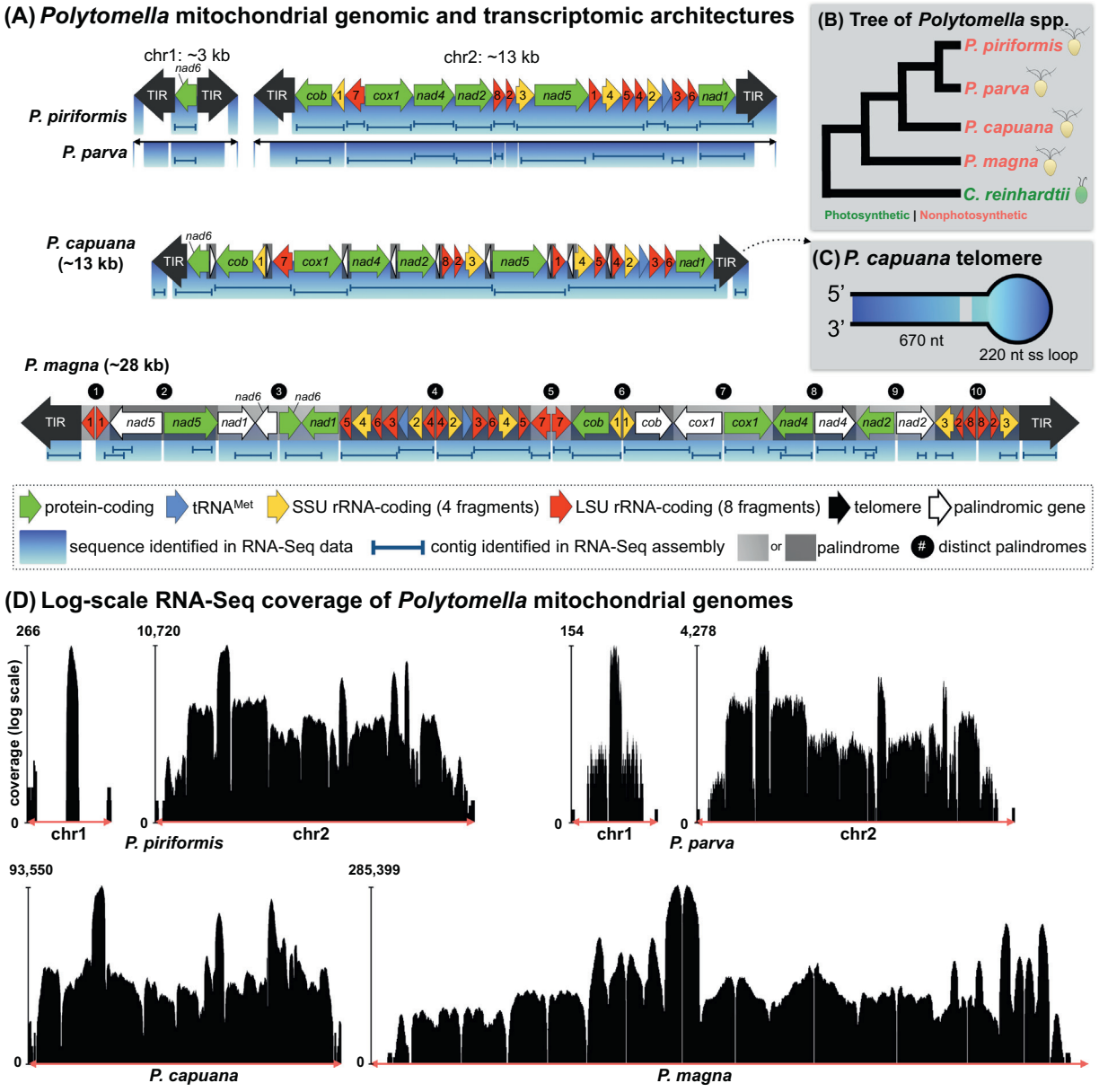

Fig. 1. *Polytomella* mitochondrial transcription. A. Mitochondrial genome maps of the four known *Polytomella* lineages. All four genomes are made up of linear chromosomes (chr) with terminal inverted repeat (TIR) telomeres and contain 10 unique genes, including the small- and large-subunit rRNA (SSU and LSU) genes, which are fragmented and scrambled into 4 and 8 loci, respectively. The *P. magna* mtDNA contains 10 palindromic repeats (boxed in dark or light gray, and labeled with black circles), which contain putative functional (green) and putative nonfunctional gene copies (white). Regions represented in the RNA-Seq data are highlighted in blue. Mitochondrial contigs identified from *de novo* transcriptome assemblies are shown with solid lines. B. Tree of *Polytomella* algae and *Chlamydomonas reinhardtii*, based on phylogenetic analysis of Smith et al. (2013). C. Transcription of single-stranded (ss) hairpin-loop telomeres from *P. capuana*. D. Log-scale RNA-Seq coverage of *Polytomella* mitochondrial chromosomes: log (coverage + 1)/log(maximum coverage + 1).

Table 1
Polytomella Illumina RNA-Seq data.

	P. piriformis	*P. parva*	*P. capuana*	*P. magna*
RNA-SEQ DATA[a]				
Number of reads	6.97×10^7	5.29×10^7	9.31×10^7	26.0×10^7
GenBank accession	SRX710732	SRX551283	SRX710731	SRX710730
MAPPING TO MITO GENOME[b]				
Reads mapped	528 | 52,790	981 | 50,968	0.32×10^6	2.2×10^6
Mean coverage (reads/nt)	23 | 556	16 | 197	3881	9939
Genome coverage (%)	55.5 | 97.2	67.6 | 90.9	99.2	95.1

[a] Paired-end Illumina RNA sequencing data (see Methods for details).
[b] Mitochondrial genome (Mito). The *P. piriformis* and *P. parva* mitochondrial genomes are comprised of two different chromosomes (Fig. 1A); a vertical line "|" separates the mapping data for chromosomes 1 and 2.

genomic research. It is perhaps best employed in conjunction with other RNA profiling techniques, and could be particularly useful for generating mitochondrial gene or genome sequences from taxa for which there is an absence of available mtDNA data.

2. Materials and methods

P. piriformis, *P. parva*, *P. capuana*, and *P. magna* (Göttingen Culture Collection of Algae, SAG, strains 63-10, 63-3, 63-5, and 63-9, respectively) were grown and harvested as previously described (Smith and Lee, 2014; MacDonald and Lee, 2015). *P. parva* RNA extraction, library preparation, Illumina sequencing, and *de novo* RNA-Seq assembly were carried out by the National Center for Genome Resources (NCGR) following the protocols of the MMETSP (Smith and Lee, 2014; Keeling et al., 2014). For *P. piriformis*, *P. capuana* and *P. magna*, total cellular RNA was isolated using the Qiagen (MD, USA) RNeasy Plant Mini Kit and treated with Qiagen RNase-free DNase followed by RNA-Seq library preparation and Illumina (HiSeq 2500) sequencing (paired end, 2×150 cycle run) at the McGill University and Genome Quebec Innovation Centre (MUGQIC). The raw RNA-Seq data were trimmed and clipped with Trimmomatic (Bolger et al., 2014) and the normalized reads were assembled with Trinity (Haas et al., 2013) by the MUGQIC Bioinformatics team using their standard parameters.

RNA-Seq reads (Table 1) were mapped to their respective *Polytomella* mitochondrial genome sequences, which have no introns, with Bowtie 2 (Langmead and Salzberg, 2012), implemented through Geneious v8.1.4 (Biomatters Ltd., Auckland, NZ) using default settings, the lowest sensitivity option, and a min/max insert size of 50 nt/500 nt. Because of the telomeric inverted repeats and palindromic genes within *Polytomella* mtDNA, each read was allowed to map to a maximum of two locations. Mitochondrial contigs were mined from *de novo* transcriptome assemblies with BlastN (Altschul et al., 1990) using entire *Polytomella* mtDNA sequences as search queries and an expectation value of $\leqslant 1e^{-10}$. The RNA-Seq data used in this study are in the NCBI Sequence Read Archive under accession numbers SRX710732 (*P. piriformis*), SRX551283 (*P. parva*), SRX710731 (*P. capuana*), and SRX710730 (*P. magna*).

3. Results and discussion

3.1. Assembling Polytomella mitochondrial genomes from RNA-Seq data

As part of a collaborative initiative to characterize non-photosynthetic green algae (Smith and Lee, 2014; MacDonald and Lee, 2015), we generated paired-end Illumina RNA-Seq data from the four known *Polytomella* species (Table 1; Fig. 1B). These data, which were derived from whole cellular RNA, contained a large number of mitochondrial-derived reads (Table 1). But as expected, no plastid reads were detected, which is consistent with the *Polytomella* plastid not having a genome or gene expression system (Smith and Lee, 2014). Mapping of the RNA-Seq data to their respective *Polytomella* mtDNAs gave deep, near-complete coverage of the genomes (Table 1; Fig. 1A and D). For instance, 99.2% of the 12,998 nt *P. capuana* mitochondrial genome was represented in the sequencing reads, leaving only 102 nt unaccounted for (all of which were from the telomeres), and implying that the bulk of the mtDNA, including all of the intergenic regions, is transcribed. Similar trends were observed for *P. piriformis*, *P. parva*, and *P. magna*, although for the latter two species small segments of intergenic as well as telomeric mtDNA were absent from the RNA-Seq data (Fig. 1A and D).

The read coverage varied within and among the different *Polytomella* mitochondrial chromosomes (Fig. 1D), but on average it was moderate to high, ranging from 16 (chromosome 1 of *P. parva*) to 9939 (*P. magna*) reads per nt. For all four mtDNAs, coverage was lowest in the intergenic and telomeric regions and highest in areas encoding rRNAs, likely reflecting their high levels of expression. Again, in some instances there were gaps in the coverage, particularly in the mitochondrial chromosome 1 telomeres of both *P. piriformis* and *P. parva* and the inverted-repeat junctions of *P. magna* (discussed below). Not surprisingly, there was a positive relationship between the size of the RNA-Seq dataset and the mean read coverage of the mitochondrial genome (Table 1)—the species for which we had generated the largest number of reads also had the highest average mitochondrial read coverage (*P. magna*).

In addition to mapping the reads onto the mitochondrial chromosomes, we also scanned *de novo* assemblies of the RNA-Seq data for mitochondrial contigs. As with the mapping experiments, large multi-gene segments of the *Polytomella* mitochondrial genomes could be recovered from the assembled reads (Fig. 1A). For instance, in the *P. capuana* transcriptome we found 9 (mostly polycistronic) mitochondrial contigs, ranging from 220 to 3320 nt, containing 1–8 coding regions apiece, and encompassing ~97.5% of the genome. Likewise, analysis of the *P. piriformis* transcriptome exposed complete transcripts for all of the defined mitochondrial genes (some of which were arranged as polycistrons), every intergenic region, and large sections of the telomeres (Fig. 1A). The *P. parva* and *P. magna* mitochondrial genomes were also well-represented in their respective transcriptome assemblies, but sequences for some mitochondrial genes, including portions of the large-subunit rRNA gene as well as various noncoding regions were incomplete or missing altogether (Fig. 1A). Overall, we were able to assemble a greater fraction of the *Polytomella* mitochondrial genomes by mapping reads than by mining contigs from *de novo* assemblies. Nevertheless, detailed inspection of the read-mapping and contig results revealed interesting and unexpected patterns of mitochondrial transcription, particularly with respect to the telomeres.

3.2. Active transcription of Polytomella mitochondrial telomeres

Although often thought of as circular molecules, many mitochondrial chromosomes are linear and have telomeres, which likely help to preserve the chromosome ends, independent of telomerase (Nosek and Tomáska, 2003). The various functions of mitochondrial telomeres, however, are not yet well understood, particularly with respect to transcription. Thus, it is intriguing that for all four *Polytomella* species large pieces of the mitochondrial telomeres could be assembled from the RNA-Seq data, indicating that these regions are actively transcribed and potentially involved in gene expression (Fig. 1A and C). These findings are supported by earlier reverse-transcriptase PCR analyses, which successfully amplified telomeric mtRNA from *Polytomella* taxa (Smith et al., 2010).

Polytomella telomeres have a complex architecture. They are ~900–1300 nt long and organized into inverted repeats whereby the sequence of one terminus is the reverse complement of the other terminus. Within species the telomeric sequences are almost identical, even between different mitochondrial chromosomes, but among species they differ substantially (Smith et al., 2010). The extreme ends of the telomeres, which have been sequenced in *P. parva*, *P. piriformis*, and *P. capuana* (but not in *P. magna*), form covalently closed hairpin loops with lengths of ~150–220 nt, depending on the species (Fig. 1C).

When looking across the different *Polytomella* telomeres, there were no obvious patterns in the breadth and depth of the RNA-Seq coverage, apart from it being low compared to coding regions and incomplete in places. Mapping the reads to the termini was complicated by the inverted-repeat nature of the telomeres, making it hard to know to which telomere the reads belonged, with the exception of reads bordering the sub-telomeric regions and/or those with a paired read anchored in non-repetitive DNA, which were straightforward to assemble. Fortunately, for both *P. parva* and *P. piriformis* the telomeric sequences of chromosomes 1 versus 2 are slightly different from one another allowing for moderately accurate telomeric read mapping between chromosomes. Complications aside, there is no doubt that in each of the *Polytomella* species we investigated a significant fraction of the mitochondrial telomeres are expressed. Moreover, in the three species for which entire telomeric sequences are available, RNA-Seq reads mapped to the very ends of the genomes; in other words, the single-stranded telomeric loops are transcribed in *Polytomella* algae (Fig. 1C).

Hairpin telomeres are found in a variety of organisms and genetic compartments, including the mitochondrion of *Paramecium* (Nosek and Tomáska, 2003) and the bacterium *Borrelia burgdorferi* (Chaconas, 2005), which causes Lyme disease. Although there has been some progress in understanding how hairpin telomeres are replicated and maintained (Nosek and Tomáska, 2003; K. Shi et al., 2013), there is little information on their transcriptional properties, especially those from mitochondrial genomes. To the best of our knowledge, the present study is one of only a few clearly showing the expression of hairpin telomeric elements. It remains to be determined whether the transcription of these elements reflects a specific regulatory role in gene expression and/or genome replication (e.g., RNA primers) or is merely transcriptional noise. One possibility, as described in the proceeding section, is that the *Polytomella* telomeres are involved in the production, cleavage, and/or processing of polycistronic mitochondrial transcripts.

3.3. Polycistronic mitochondrial transcripts and the expression of palindromic genes

The results from both the RNA-Seq mappings and *de novo* transcriptome assemblies suggest that *Polytomella* mitochondrial genes are, with some exceptions, expressed as polycistronic transcripts, which are then cleaved and processed into smaller polycistronic and monocistronic units (Fig. 1A). A polycistronic organization is supported by the recovery of RNA contigs containing multiple mitochondrial genes as well as RNA-Seq coverage of intergenic regions, including reads situated entirely within or spanning whole intergenic spacers and paired reads anchored in neighboring genes (e.g., one read in *nad4* and its partnered read in *nad2*) (Fig. 1A). However, the sharp drop in read coverage at intergenic spacers and telomeres relative to coding regions suggests that large polycistronic transcripts are efficiently cleaved and processed at noncoding sites into smaller pieces and eventually individual coding units, which represent the majority of the RNA species within the mitochondria (Fig. 1D), as previously shown for *P. parva* (Fan et al., 2003) and *C. reinhardtii* (Gray and Boer, 1988).

The size, organization, and content of the mitochondrial polycistrons likely vary among the different *Polytomella* species. For example, in *P. capuana* the mapping data support the presence of a genome-sized polycistronic transcript containing portions of both telomeres (Fig. 1A). Because the *P. capuana* mtDNA is organized into two opposing transcriptional units, a genome-length transcript would need to be generated for each strand of the mtDNA. Deep RNA-Seq coverage spanning the junction between the two *P. capuana* transcriptional units supports the presence of such transcripts (Fig. 1D). Likewise, the *P. parva* and *P. piriformis* mtDNAs (Fig. 1D) also appear to be expressed as chromosome-sized units, and for both taxa RNA-Seq coverage was high in locations of opposing transcriptional polarity (i.e., the *cox1–L7* intergenic spacer) (Fig. 1D).

Unraveling the mitochondrial expression patterns of *P. magna* was more challenging than for the other three *Polytomella* species. The *P. magna* mtDNA is made up of ten large inverted repeats containing 1–7 coding segments apiece (Fig. 1A) (Smith et al., 2013). Consequently, every gene is present twice in the genome. There can be, however, small differences between duplicate protein-coding genes, such as premature or missing start/stop codons, implying that only one of the gene copies is functional (Smith et al., 2013). Mapping of the RNA-Seq data onto the *P. magna* mtDNA gave some insights into the transcription of the duplicate genes, but not as many insights as we had hoped for.

Like with the telomeres, it was hard to decipher which reads belonged to which copy of the *P. magna* inverted repeats, and because of this we allowed each read to map to the reference genome up to two times. We paid particularly close attention to the read mapping in the intersections between different inverted repeat elements (e.g., the *nad5-nad1* intergenic spacer) and between identical inverted repeat pairs (e.g., the *nad5-nad5* intergenic spacer) (Fig. 1A), both of which contain short stretches of non-repetitive sequence. The RNA-Seq data mapped to and spanned the areas between different inverted repeats, but only when the intersections contained a shift in transcriptional polarity, such as the *L1-nad5* intergenic spacer (Fig. 1A). That is to say, no reads or contigs mapped to inverted-repeat intersections in which the direction of transcription stayed the same. Similarly, not a single read or contig spanned the border between the inverted-repeat pairs (i.e., the region representing the hairpin loop in a secondary structure diagram of the repeat element). This lack of coverage at the intersections between identical and distinct inverted repeat elements may reflect highly efficient post- or co-transcriptional cleavage and processing at these sites, or possibly that these regions are not transcribed at all.

Unfortunately, the absence of read mapping in the inverted-repeat border regions prevented us from reliably identifying which copies of the duplicate genes are transcribed and "functional". In some cases, the read coverage indicated that both gene copies (or at least portions of both), including those of *nad2*, *nad4*, *cox1*,

and *cob*, are transcriptionally active, leaving open the possibility that each copy of the duplicate gene gives rise to a functional sequence. However, the RNA-Seq data alone cannot be used to accurately test such a hypothesis. Altogether, the *P. magna* analyses raised more questions than they answered, and although the data were intriguing, they also exposed some of the limitations of an RNA-Seq-only approach to mitochondrial transcriptomics: RNA read-mapping alone is often not enough to identify transcript boundaries within complex, repeat-rich organelle genomes.

3.4. Pros and cons of RNA-Seq for mitochondrial genomics

A major goal of this study was to evaluate the utility of RNA-Seq for recovering mitochondrial genome sequences. On this front, we were successful: moderate amounts of RNA-Seq data from *Polytomella* species easily gave near-complete mitochondrial genome assemblies. The extremely small sizes and polycistronic transcriptional organizations of *Polytomella* mitochondrial chromosomes undoubtedly facilitated their efficient recovery from the RNA data. Nonetheless, we believe that the approach employed here can be used to generate organelle genome sequences from other eukaryotic species, including those with larger mtDNAs than *Polytomella* spp.

It is becoming more and more apparent that polycistronic transcription is a common theme among mitochondria (Barbrook et al., 2010), with some notable exceptions (Lukeš et al., 2005; Marande and Burger, 2007). From humans (Mercer et al., 2011) to slime molds (Le et al., 2009) to malaria parasites (Rehkopf et al., 2000) to green algae (Gray and Boer, 1988), mtDNAs are usually expressed as long multi-gene precursor transcripts, which include noncoding regions. Take, for example, the 56 kb mtDNA of *Dictyostelium discoideum*, which is transcribed from a single unidirectional promoter and gives rise to a near-genome-sized polycistron (Le et al., 2009), or the *C. reinhardtii* mtDNA, which is expressed by the generation of long, multi-gene co-transcripts (Gray and Boer, 1988; Wobbe and Nixon, 2013), similar to those proposed above for its close *Polytomella* relatives. Large polycistronic precursor RNAs have also been observed in the mitochondria from a diversity of other eukaryotes (Barbrook et al., 2010), including *Tetrahymena pyriformis* and *Chondrus crispus* (Richard et al., 1999; Edqvist et al., 2000), to name but two, as well as in a wide range of plastids (Stern et al., 2010; C. Shi et al., 2013).

The proclivity for polycistronic transcription within mitochondria and plastids, means that RNA-Seq can (and should) be used to assemble and analyze large sections of organelle chromosomes, as outlined here for *Polytomella* [and previously for the plastid genomes of *Camellia* (C. Shi et al., 2013)]. And now that online genetic databanks are swelling with eukaryotic RNA-Seq reads, the time is ripe for exactly these types of analyses. That said, surprisingly few researchers are using the available RNA-Seq data for gleaning organelle sequences. As well as providing information on mitochondrial chromosome and transcriptional architecture, RNA-Seq data can be mined for specific mitochondrial gene sequences or loci, facilitating phylogenetic, population genetic, comparative genomic, and genetic barcoding analyses. Moreover, RNA-Seq is an excellent tool for examining the frequent and widespread post-transcriptional editing found in various protist and land plant mitochondria (Smith and Keeling, 2015), and for lineages with severe forms of post-transcriptional editing, such as angiosperms, RNA-Seq assemblies might in fact provide more coherent and "useable" genetic information than the mtDNA sequences alone. Indeed, RNA-Seq could be especially useful for studying organisms that have not yet had their mtDNAs sequenced but for which RNA-Seq data are available. A scan of the organisms selected for the Marine Microbial Eukaryotic Transcriptome Sequencing Project reveals dozens of such cases (Keeling et al., 2014).

While providing an easy avenue for generating organelle sequences and exploring transcription, RNA-Seq is not without its drawbacks, some of which were encountered here in our investigations of *Polytomella*. First, there is always the potential for mistaking nuclear-genome-encoded mitochondrial-like transcripts for genuine mtRNA (Kleine et al., 2009), which is more of a problem in land plants and animals than in protists (Smith et al., 2011). Second, species with monocistronic mitochondrial transcription units, such as kinetoplastids and diplonemids (Lukeš et al., 2005; Marande and Burger, 2007), may greatly limit the amount mitochondrial genomic information that can be mined from RNA-Seq reads, but such species are the exception rather than the norm. And last, RNA-Seq information by itself is sometimes not enough to accurately characterize the transcriptional landscapes of complex mitochondrial genomes, like that of *P. magna* and other protists. Indeed, with standard RNA-Seq techniques it is often not possible to distinguish between sense and antisense transcripts or to precisely identify transcript cleavage sites. Ultimately, for a detailed understanding of mitochondrial transcription, RNA-Seq is best used alongside other RNA profiling methods, such as RNA electrophoresis, Northern blotting, quantitative real-time reverse-transcription PCR, and/or directional RNA sequencing. By itself, RNA-Seq is an excellent but under-used technique for recovering mitochondrial gene or genome sequences. The databanks are waiting, so let's start mining.

Acknowledgments

We thank Jimeng Hua and Robert W. Lee for generating and providing much of the RNA-Seq sequence data that were used in this study. The latter also gave helpful feedback on the manuscript. This work was supported by a Discovery Grant to DRS from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Barbrook, A.C., Howe, C.J., Kurniawan, D.P., Tarr, S.J., 2010. Organization and expression of organellar genomes. Philos. Trans. Roy. Soc. B 365, 785–797.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, btu170.
Chaconas, G., 2005. Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end. Mol. Microbiol. 58, 625–635.
Edqvist, J., Burger, G., Gray, M.W., 2000. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. J. Mol. Biol. 297, 381–393.
Fan, J., Lee, R.W., 2002. Mitochondrial genome of the colorless green alga *Polytomella parva*: two linear DNA molecules with homologous inverted repeat termini. Mol. Biol. Evol. 19, 999–1007.
Fan, J., Schnare, M.N., Lee, R.W., 2003. Characterization of fragmented mitochondrial ribosomal RNAs of the colorless green alga *Polytomella parva*. Nucleic Acids Res. 31, 769–778.
Gray, M.W., Boer, P.H., 1988. Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial-DNA. Philos. Trans. Roy. Soc. Lond. Ser. B: Biol. Sci. 319, 135–147.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., et al., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.
Hodgkinson, A., Idaghdour, Y., Gbeha, E., Grenier, J.C., Hip-Ki, E., et al., 2014. High-resolution genomic analysis of human mitochondrial RNA sequence variation. Science 344, 413–415.
Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., et al., 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12, e1001889.
Kleine, T., Maier, U.G., Leister, D., 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Ann. Rev. Plant Biol. 60, 115–138.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Le, P., Fisher, P.R., Barth, C., 2009. Transcription of the *Dictyostelium discoideum* mitochondrial genome occurs from a single initiation site. RNA 15, 2321–2330.

Lukeš, J., Hashimi, H., Zíková, A., 2005. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. Curr. Genet. 48, 277–299.
MacDonald, S.M., Lee, R.W., 2015. Validation of *Polytomella piriformis* nomen nudum (Chlamydomonadaceae): a distinct lineage within a genus of nonphotosynthetic green algae. J. Euk. Microbiol. 62, 840–844.
Marande, W., Burger, G., 2007. Mitochondrial DNA as a genomic jigsaw puzzle. Science 318, 415.
Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., et al., 2011. The human mitochondrial transcriptome. Cell 146, 645–658.
Metzker, M.L., 2010. Sequencing technologies – the next generation. Nat. Rev. Genet. 11, 31–46.
Nakada, T., Misawa, K., Nozaki, H., 2008. Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. Mol. Phylogenet. Evol. 48, 281–291.
Nosek, J., Tomáska, L., 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. Curr. Genet. 44, 73–84.
Rackham, O., Shearwood, A.M.J., Mercer, T.R., Davies, S.M., Mattick, J.S., et al., 2011. Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. RNA 17, 2085–2093.
Raz, T., Kapranov, P., Lipson, D., Letovsky, S., Milos, P.M., et al., 2011. Protocol dependence of sequencing-based gene expression measurements. PLoS ONE 6, e19287.
Rehkopf, D.H., Gillespie, D.E., Harrell, M.I., Feagin, J.E., 2000. Transcriptional mapping and RNA processing of the *Plasmodium falciparum* mitochondrial mRNAs. Mol. Biochem. Parasitol. 105, 91–103.
Richard, O., Kloareg, B., Boyen, C., 1999. MRNA expression in mitochondria of the red alga *Chondrus crispus* requires a unique RNA-processing mechanism, internal cleavage of upstream tRNAs at pyrimidine 48. J. Mol. Biol. 288, 579–584.
Shi, C., Liu, Y., Huang, H., Xia, E.H., Zhang, H.B., et al., 2013. Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of ycf15 function and evolution in angiosperms. PLoS ONE 8, e59620.
Shi, K., Huang, W.M., Aihara, H., 2013. An enzyme-catalyzed multistep DNA refolding mechanism in hairpin telomere formation. PLoS Biol. 11, e1001472.
Smith, D.R., 2012. Not seeing the genomes for the DNA. Brief. Funct. Genom. 11, 289–290.
Smith, D.R., 2013. RNA-Seq data: a goldmine for organelle research. Brief. Funct. Genom. 12, 454–456.
Smith, D.R., 2016. The past, present, and future of mitochondrial genomics: have we sequenced enough mtDNAs? Brief. Funct. Genomics. 15, 47–54.
Smith, D.R., Crosby, K., Lee, R.W., 2011. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. Genome Biol. Evol. 3, 365–371.
Smith, D.R., Hua, J., Archibald, J.M., Lee, R.W., 2013. Palindromic genes in the linear mitochondrial genome of the nonphotosynthetic green alga *Polytomella magna*. Genome Biol. Evol. 5, 1661–1667.
Smith, D.R., Hua, J., Lee, R.W., 2010. Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. Curr. Genet. 56, 427–438.
Smith, D.R., Keeling, P.J., 2015. Mitochondrial and plastid genome architecture: reoccurring themes but significant differences at the extremes. Proc. Natl. Acad. Sci. U.S.A. 112, 10177–10184.
Smith, D.R., Lee, R.W., 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. Mol. Biol. Evol. 25, 487–496.
Smith, D.R., Lee, R.W., 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Phys. 164, 1812–1819.
Stern, D.B., Goldschmidt-Clermont, M., Hanson, M.R., 2010. Chloroplast RNA metabolism. Ann. Rev. Plant Biol. 61, 125–155.
Wobbe, L., Nixon, P.J., 2013. The mTERF protein MOC1 terminates mitochondrial DNA transcription in the unicellular green alga *Chlamydomonas reinhardtii*. Nucleic Acids Res. 41, 6553–6567.