Received: 4 April 2024 | Accepted: 28 May 2024

DOI: 10.1111/ppl.14418

VIEWPOINT

Physiologia Plantarum

Long-read RNA-Seq for the discovery of long noncoding and antisense RNAs in plant organelles

Matheus Sanita Lima[1] | Douglas Silva Domingues[2] | Alexandre Rossi Paschoal[3] | David Roy Smith[1]

[1]Department of Biology, Western University, London, ON, Canada

[2]Department of Genetics, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, SP, Brazil

[3]Department of Computer Science, Bioinformatics and Pattern Recognition Group (BIOINFO-CP), Federal University of Technology–Paraná–UTFPR, Cornélio Procópio, PR, Brazil

Correspondence
Matheus Sanita Lima
Email: msanital@uwo.ca

David Roy Smith
Email: dsmit242@uwo.ca

Funding information
Fundação Araucária, Grant/Award Number: NAPI Bioinformatica 66.2021; Natural Sciences and Engineering Research Council of Canada, Discovery Grant

Edited by J.-.F. Mao

Abstract

Plant organelle transcription has been studied for decades. As techniques advanced, so did the fields of mitochondrial and plastid transcriptomics. The current view is that organelle genomes are pervasively transcribed, irrespective of their size, content, structure, and taxonomic origin. However, little is known about the nature of organelle noncoding transcriptomes, including pervasively transcribed noncoding RNAs (ncRNAs). Next-generation sequencing data have uncovered small ncRNAs in the organelles of plants and other organisms, but long ncRNAs remain poorly understood. Here, we argue that publicly available third-generation long-read RNA sequencing data from plants can provide a fine-tuned picture of long ncRNAs within organelles. Indeed, given their bloated architectures, plant mitochondrial genomes are well suited for studying pervasive transcription of ncRNAs. Ultimately, we hope to showcase this new avenue of plant research while also underlining the limitations of the proposed approach.

1 | INTRODUCTION AND RESULTS

When it comes to extremes in organelle genome transcription, plants continue to outdo themselves. Whether it is unusual modes of trans-splicing, huge numbers of RNA-editing sites, or an overabundance of introns, you name it, there is a species that has done it (Sloan et al., 2012; Rice et al., 2013; Su et al., 2019). These eccentricities have been studied via laborious benchtop work (Sanita Lima et al., 2016) as well as through next generation sequencing (NGS) (Forsythe et al., 2022). Second-generation (short-read) RNA-Seq data, in particular, have been useful in documenting genome-wide, pervasive transcription of organelle genomes across the eukaryotic domain (Sanita Lima & Smith, 2017a,b). RNA-Seq has also helped elucidate, for example, multiple transcription start sites (TSS) across plastome (i.e., chloroplast genome) operons in barley (Zhelyazkova et al., 2012) and dozens of small antisense RNAs (asRNAs) in the *Arabidopsis* chloroplast genome (Hotto et al., 2011). More than anything, NGS data have highlighted the prevalence and abundance of noncoding RNAs (ncRNAs) in organelle systems (Dietrich et al., 2015), building upon earlier pioneering work from the 1990s (Vera & Sugiura, 1994).

Nevertheless, we still have a lot to learn about plant organelle noncoding transcriptomes, particularly long ncRNAs, which could have potential functions inside and/or outside of organelles (Sanita Lima et al., 2024a,b). Thankfully, publicly available third-generation (long-read) RNA-Seq data provide an unprecedented opportunity for elucidating long ncRNAs from plant organelle genomes. Indeed, here, we show that plant organelles produce kilobase-long noncoding and antisense RNAs, some of which may be functional. Our

analyses are preliminary but already point to the utility of mining whole-cell/bulk tissue long RNA reads for the study of plant organelle noncoding transcriptomes. Our initial investigations show not only promising prospects but also challenges pertaining to (meta) data annotation and interpretation. In this Viewpoint, we share the ups and downs of using PacBio and Oxford Nanopore Technologies (ONT) RNA reads to produce, what we call, "third generation" plant organelle transcriptomes.

Clues that some plants might be producing a myriad of asRNA and ncRNAs across their entire organelle genomes came from broadscale Illumina RNA-Seq mapping studies showing complete coverage of plastomes and mitogenomes (Sanita Lima & Smith, 2017a,b). These findings were, at times, hard to interpret because the repeat-rich nature of plant organelle DNA meant that a single Illumina read could map to multiple locations in a genome. Moreover, genome-wide high RNA-Seq coverage made it difficult to dissect transcript isoform architecture, including whether transcripts were long or ultra-long. PacBio and ONT long RNA reads have alleviated many of these problems (Kovaka et al., 2023; Zhou et al., 2023). Not only does long RNA-Seq deliver more data per read, but the read itself can also provide insights into transcriptional architecture, often representing a complete transcript (mature or not). Long reads are especially helpful for examining plant organelle ncRNAs because they can span entire intergenic regions, even in the largest of genomes, leaving little doubt about mis-mappings due to repeats. Even though current long-read RNA-Seq technologies are not readily strand-specific (with the exception of ONT direct-RNA-Seq) (Schuster et al., 2023), when multiple genes with opposing transcriptional polarities are located on a single long read, one can glean information about putative asRNA (Figures 1 and 2). Strand-specific Illumina reads have already demonstrated the pervasive transcription of small antisense RNAs in organelles (Hotto et al., 2011; Mercer et al., 2011), and these datasets can be combined with third-generation data to create strand-specified long reads as well (Ghareghani et al., 2018).

Long reads still represent a small fraction (<1%) of the existing RNA-Seq data in NCBI's Sequence Read Archive (SRA), but PacBio and ONT platforms are increasing in popularity, particularly within plant model systems. By cross-referencing available datasets with plant species that have fully sequenced organelle genomes, we identified 450 candidate taxa for long-read analysis (Table S1). Straightforward long-read RNA-mapping analyses (Materials and methods) show a prevalence of long antisense and noncoding transcripts across plant organelle genomes.

For instance, five ONT RNA-Seq datasets, comprising a total of 5.7 million reads from two different studies, mapped the entire plastid genome of *Solanum lycopersicum* (tomato), which is about 155 kilobases (kb) in size, with >1,300-fold coverage. 2,933 of the mapped reads exceeded 5 kb, providing a fine-tuned snapshot of the tomato chloroplast transcriptome (Table S2). In numerous regions, the reads spanned sets of genes with opposing transcriptional polarities (Figure 1). Upon visual inspection of the mapping results, we observed at least 44 instances of plastid genes being transcribed in the antisense direction, representing >28 kb of asRNA. Even more impressive, each of the ~132 intergenic regions (with one exception) could be independently captured by a single read. Near identical trends were identified in the plastomes of *Triticum aestivum* and *Aegilops tauschii* as well as the red alga *Cyanidioschyzon merolae* (Table S3), suggesting that this might be a recurring theme not only for plants but for plastid-bearing eukaryotes more generally.

As with the plastome, the entire tomato mitogenome (446 kb) was covered by long RNA reads (mean coverage = 1,186). We found multiple instances of putative asRNAs (Figure 2), but the most conspicuous observation was the preponderance of long ncRNAs. The *S. lycopersicum* mitogenome contains ~414 kb of noncoding DNA, most of which comes from its ~404 kb of intergenic nucleotides (~68 intergenic spacers with an average length of 5.9 kb - Table S4; Materials and Methods). 36 of those intergenic spacers were covered by uninterrupted long reads, indicating that the tomato mitochondrion is generating hundreds of kilobases of long ncRNAs (Table S2). Our analyses of 26 other plant (and one red alga) mitochondrial transcriptomes gave similar conclusions (Table S5).

The length distribution of mapped reads across the species we analyzed (Table S2) indicates that there is only a small difference in expression between coding and noncoding segments of plant organelle genomes. The abundance of >5 kb transcripts is also evidence for the ubiquity of polycistronic transcripts in plant organelles, which must rely heavily on posttranscriptional processing to produce gene-size transcripts (Stern et al. 2010). In certain cases (*e.g.*, the *Citrus unshiu* mitogenome), the most commonly mapped reads were almost 5 kb long (Table S2), reinforcing the idea that long RNAs can be the norm rather than the exception in plant organelles.

Given the widespread occurrence of long ncRNAs in plant organelles, especially mitochondria (Tables S3, S5), one must ask: how much of it is merely transcriptional noise? Moreover, if long noncoding segments are, indeed, generated via readthrough transcription, what consequences does the accumulation and degradation of these transcripts have on organelle homeostasis? The "reality of pervasive transcription" has been vehemently defended and contested in nuclear systems (Clark et al., 2011; van Bakel et al., 2011), where noncoding transcripts have been referred to as dark matter RNA (van Bakel et al., 2010). Ultimately, the burden of proof lies on the ncRNAs themselves, which may not be a bad thing for the field of plant organelle transcriptomics. One asset of studying organelle transcriptomes in plants is the presence of widespread RNA editing, which can be used to flag potential biological functions in noncoding transcripts (Wu et al., 2015). Also, with the ongoing development of novel organelle genome manipulation techniques, targeted mutation experiments could easily be used to probe the importance of organelle noncoding nucleotides (Wang et al., 2024).

Another advantage of plant organelle RNA editing for studying transcriptomics is that the editing patterns can be used to identify possible DNA contamination in the mapped datasets. We did not systematically evaluate RNA editing in our mapping analyses, but future studies can use the percentage of edited mapped transcripts as a proxy

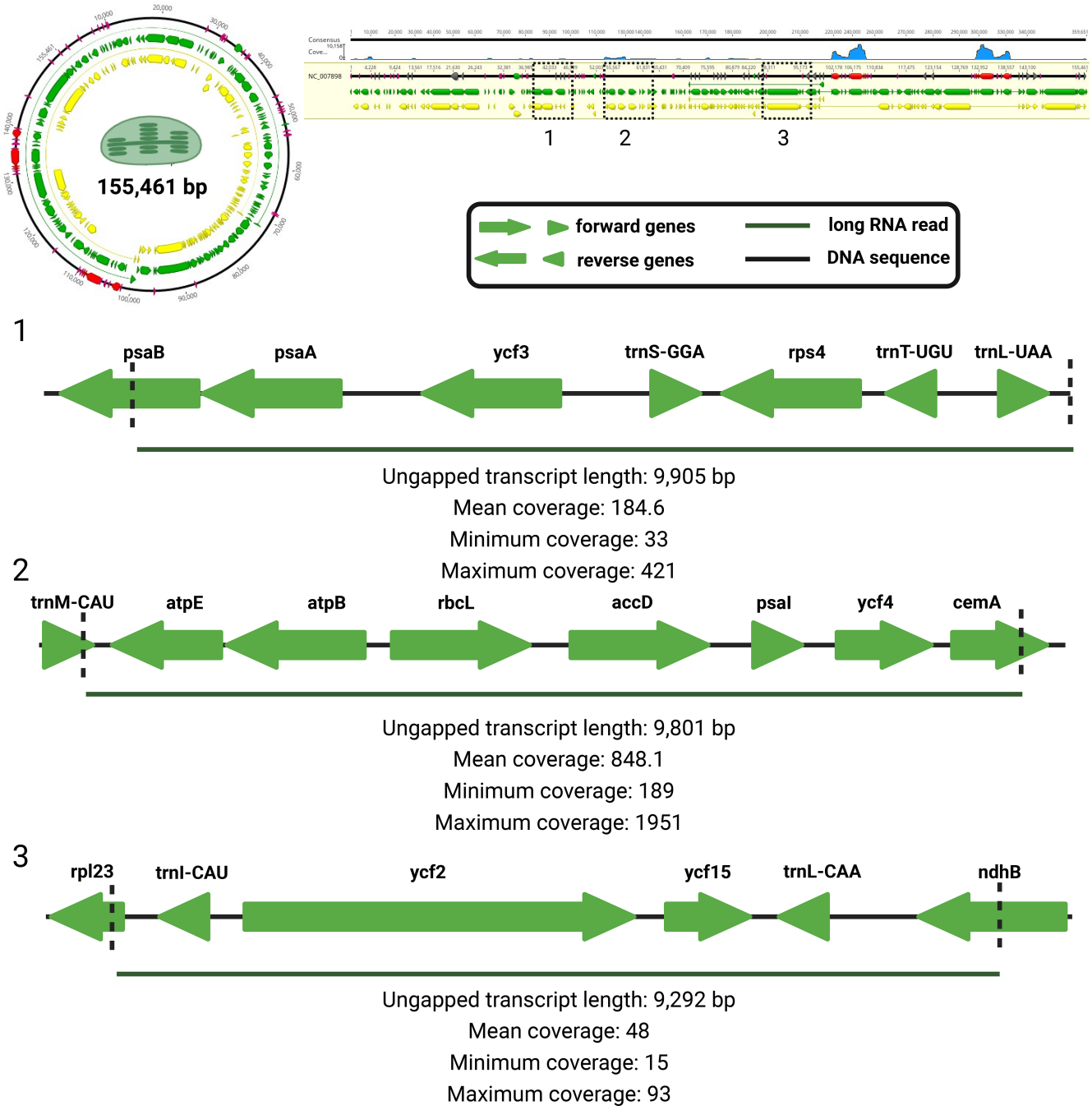

FIGURE 1 Antisense RNAs (asRNAs) from the tomato (*Solanum lycopersicum*) chloroplast genome. The circular map of the tomato plastome (NC_007898.3; 155,461 bp) is shown in the top left. The RNA-mapping histogram with coverage depth by nucleotide on the linearized molecule is shown in the top right. Three examples of asRNAs are highlighted by the dashed boxes (1, 2, and 3), and each example is depicted in greater detail below the histogram. These reads were the three longest reads mapped onto the tomato plastome. Green arrows illustrate genes and their respective polarity. The black bar represents the genome and the corresponding intergenic regions. The green bar represents a long RNA read produced from the region highlighted. Diagrams are not drawn to scale, but the length of the arrows are meant to indicate the relative size of the corresponding genes. The abbreviated gene names are listed above each gene depiction. Read mapping statistics for the specific region covered by each long RNA read are listed below each diagram. Mean, minimum, and maximum coverage indicates the number of RNA reads mapped onto the highlighted region. The three chosen long reads matched the respective genome sequence with 100% similarity. The circular map, the mapping histogram, and the mapping statistics were obtained from Geneious Prime 2023.2.1 (https://www.geneious.com). Figure created with BioRender.com.

for the purity of the RNA datasets. Should most reads exhibit edited nucleotides, the dataset(s) could be assumed to be free of DNA contamination. As plant organelle genomes are rich in introns (Mukhopadhyay & Hausner, 2021), those could also be used as natural spikes (for the presence of DNA contamination), depending on the type of sequencing library protocol that was employed (*e.g.*,

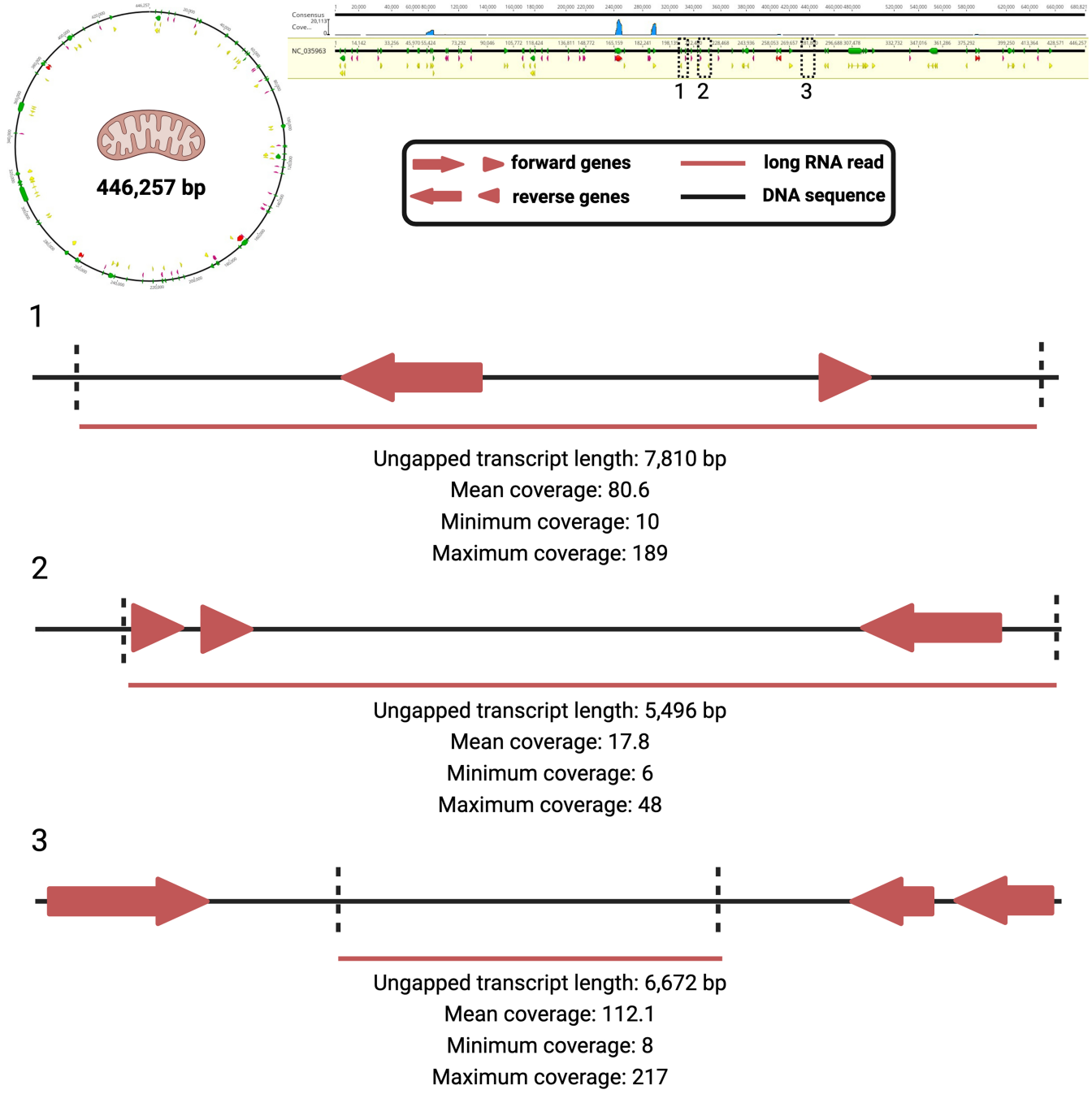

FIGURE 2 Antisense RNAs (asRNAs) and noncoding RNA (ncRNA) from the tomato (*Solanum lycopersicum*) mitochondrial genome. The circular map of the tomato mitogenome (NC_035963.1; 446,257 bp) is shown in the top left. The RNA-mapping histogram with coverage depth by nucleotide on the linearized molecule is shown in the top right. Two examples of asRNAs and one example of a ncRNA are highlighted by the dashed boxes (1, 2, and 3). Each example is depicted in greater detail below the histogram. Magenta arrows illustrate genes and their respective polarity. The black bar represents the genome and the corresponding intergenic regions. The magenta bar represents the long RNA read produced from the region highlighted. Diagrams are not drawn to scale, but the length of the arrows is meant to indicate the relative size of the corresponding genes. The abbreviated gene names are listed above each gene depiction. Read mapping statistics for the specific region covered by each long RNA read are listed below each diagram. Mean, minimum, and maximum coverage indicates the number of RNA reads mapped onto the highlighted region. The three chosen long reads matched the respective genome sequence with 100% similarity. The circular map, the mapping histogram, and the mapping statistics were obtained from Geneious Prime 2023.2.1 (https://www.geneious.com). Figure created with BioRender.com.

DNAse - vs DNAse + treatments) and the efficiency of intron-splicing. It is important to keep in mind, however, that within plant organelles there typically exists a low level of un-edited and/or intron-containing transcripts and that the presence of such (low-frequency) transcripts in RNA-seq datasets is not necessarily an indication of DNA contamination.

Despite the importance of knowing the sequencing protocols when mining publicly available datasets, our results demonstrate that different sequencing library preparation methods were not a major impediment to our approach. Because polyadenylation of plastid and mitochondrial transcripts mostly means RNA degradation in plants (Stern et al. 2010; Chang & Tong, 2012), future studies should compare whether data obtained with different sequencing library protocols (e.g., polyA-selection vs rRNA depletion) produce diverging transcriptional maps. Without strand-specific sequencing data, we were unable to say precisely which genes are in the antisense state when oriented in opposing transcriptional polarities to one another (Figure 1). Nevertheless, we can confidently state that large amounts of asRNA and ncRNA are being generated from these genomes. Third-generation sequencing technologies (ONT direct RNA-Seq notwithstanding) are not readily strand-specific (Kainth et al., 2023), but long reads can be "stranded" (*i.e.*, gain strand of origin information) when combined with strand-specific short reads (Ghareghani et al., 2018) or when processed through specific bioinformatic pipelines (such as Restrander) (Schuster et al., 2023). As of now, only a few plant species, such as *Arabidopsis* (Hotto et al., 2011), have strand-specific RNA data, so determining the true extent of long asRNAs in plant organelle genomes will have to await new datasets.

Another challenge of using whole-cell transcriptomic data is the potential presence of nuclear-mitochondrial and nuclear-plastid DNA sequences, NUMTs and NUPTs, respectively (Kleine et al., 2009). NUMTs and NUPTs are widespread in plants (*e.g.*, *Arabidopsis thaliana* has more than the entire mitogenome inserted in its nuclear chromosome 2) (Fields et al., 2022), making it challenging, at times, to distinguish between *bona fide* organelle transcripts and NUMTs or NUPTs. In vertebrates, the transcription of mitochondrial small ncRNAs is not explained by the number of NUMTs (Pozzi & Dowling, 2019), but kilobase-long NUMT fragments can reach 100% similarity with the mitogenome in *A. thaliana* (Fields et al., 2022). All in all, long and ultra-long RNA reads could span nuclear junction regions (or entire NUMTs/NUPTs) and be filtered out, depending on the overall read similarity. The assumption that pseudogene-derived transcripts (*i.e.*, NUMT/NUPT RNA reads) would exhibit a < 95% similarity when mapped to organelle genomes (Dietrich et al., 2015) could be tested with long-read RNA-Seq data as well.

Organelle-derived reads within genomic and transcriptomic datasets are typically treated as contamination (Smith, 2013), but publicly available whole-cell/bulk tissue sequencing data represent an untapped resource for organelle biology studies. Second-generation RNA-Seq data not only allow for the easy reconstruction of organelle genomes (Forni et al., 2019) that can, for example, be used for phylogenetic/barcoding purposes (Crampton-Platt et al., 2016; Kobayashi, 2023), but these data can shed new light into the complexities of organelle genome transcription (Smith & Sanita Lima, 2017). However, transcriptomic data (either short- or long-read) seem to be underutilized in the studies of organelle genomes (Sanita Lima et al., 2016). Whereas several dedicated pipelines have been created to assemble organelle genomes from short-read genomic data (*e.g.*, NOVOPlasty and Norgal) (Al-Nakeeb et al. 2017; Dierckxsens et al., 2017), only one bioinformatics tool (MITGARD) uses short RNA-Seq data as input for the assembly of mitochondrial genomes (Nachtigall et al., 2021). In the meantime, pipelines designed to use long-read DNA data (e.g., MitoHiFi and mitoVGP) have enabled the assembly of hundreds of mitochondrial genomes (Formenti et al., 2021; Uliano-Silva et al., 2023). ChloroSeq is the only tool developed to map chloroplast transcriptomes from short RNA-Seq data (Castandet et al., 2016), and to our knowledge, no similar tool has been created to employ long RNA reads in the investigations of organelle transcriptomes (however, MITGARD has recently been augmented to accept long read RNA-Seq data as input as well). Indeed, while Illumina RNA data have been used to study organelle genome transcription in barley, *Arabidopsis*, and dozens of plastid-bearing species (Hotto et al., 2011; Zhelyazkova et al., 2012; Sanita Lima & Smith, 2017a,b), long RNA reads have been employed to chart organelle transcriptomes in fewer than ten species (Guilcher et al., 2021; Sanita Lima et al., 2024b). Here, we demonstrate that third-generation, long-read RNA data are a valuable resource for organelle biologists. Our long-read RNA mapping analyses corroborate the genome-wide, pervasive transcription of plant organelle genomes and point to the existence of putative long noncoding and antisense transcripts in the respective (noncoding) transcriptomes.

There is accumulating data that antisense and noncoding transcripts play important roles in organelle function and gene expression (Georg et al., 2010; Hotto et al., 2012; Dietrich et al., 2015; Ruew et al., 2019; Anand & Pandi, 2021). Previous studies, however, have focused almost entirely on short (<1 kb) asRNAs and ncRNAs. Given the consistently high coverage of both long asRNA and ncRNA in our organelle transcriptome analyses, it is reasonable to suggest that at least some of these long transcripts might be functionally operational within plant mitochondria and chloroplasts. Organelle genetics has taught us that one needs to be careful before assigning utility to unusual features or, at the very least, arguing for adaptive origins for the evolution of such features (Gray et al., 2010). We accept that the null hypothesis should be that most noncoding transcripts do not have an immediate biological function (Palazzo & Gregory, 2014); however, we also believe that transcripts from organelle noncoding sequences can be exapted for regulatory roles (Sanita Lima et al., 2024b). From the arduous wet lab analyses of the previous century to the current massively parallel RNA sequencing technologies, plant organelle genomes have always occupied a space of prominence in genetics and related disciplines. Plant organelle transcriptomes, now in their third generation, will continue to be a testament to the versatility and complexity of plant mitochondria and chloroplasts.

2 | MATERIALS AND METHODS

2.1 | Data acquisition

PacBio and Oxford Nanopore Technologies (ONT) RNA-Seq datasets were retrieved from the NCBI Sequence Read Archive (SRA - https://www.ncbi.nlm.nih.gov/sra). All unique identifiers for each dataset (e.g., SRR8873358) are listed in the Supporting Information Tables. Datasets were downloaded to our private laboratory server for downstream analyses.

2.2 | RNA mapping

RNA mapping analyses were performed in Geneious Prime 2023.2.1 (https://www.geneious.com). Long RNA-Seq reads derived from whole-cell/bulk tissue experiments were mapped onto the corresponding organelle genomes using the built-in Geneious mapper/assembler. Mapper settings were mostly in default mode but varied depending on the size of the dataset and CPU demand. After testing for several parameters, we determined that no mapping iterations were needed, and mapping sensitivity could be kept at Medium-Low or Medium.

2.3 | Organelle genome composition (coding vs noncoding) and RNA read length distribution

The calculations of each component (coding vs noncoding nucleotides) from the studied plastid and mitochondrial genomes were performed using the original annotations from NCBI. Annotations were retrieved and sorted out via Geneious Prime 2023.2.1. The calculations provide only approximate estimates because nested genes and "hypothetical protein" annotations were not removed. Read length distribution was calculated using the read length graphs from Geneious Prime 2023.2.1.

AUTHOR CONTRIBUTIONS

DRS conceptualized the study and wrote the first version of the manuscript. MSL performed the analyses, designed Figures 1 and 2, compiled the Supporting Information, and edited the manuscript. DSD and ARP discussed results, suggested further analyses, edited the manuscript, and contributed to conclusions.

ACKNOWLEDGEMENTS

We thank Breno Dupin and Vitor Gregorio from Federal University of Technology – Paraná – (UTFPR) for always being promptly available to help with challenges pertaining to data retrieval and analysis.

FUNDING INFORMATION

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) via a Discovery Grant to DRS, and by Fundação Araucária (Project: NAPI Bioinformatica) via grant 66.2021 to ARP and DSD.

CONFLICT OF INTEREST STATEMENT

None of the authors have a conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

All data obtained through our analyzes are presented either in the main text or in the Supporting Information Tables. Publicly available RNA-Seq datasets were mined from NCBI SRA (Sequence Read Archive - https://www.ncbi.nlm.nih.gov/sra). Each dataset's unique identifier is listed in the Supporting Information Tables.

ORCID

Matheus Sanita Lima https://orcid.org/0000-0002-4050-4022
Douglas Silva Domingues https://orcid.org/0000-0002-1290-0853
Alexandre Rossi Paschoal https://orcid.org/0000-0002-8887-0582
David Roy Smith https://orcid.org/0000-0001-9560-5210

REFERENCES

Al-Nakeeb K, Petersen TN, Sicheritz-Ponten T. 2017. Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinformatics* 18: 510.

Anand A, Pandi G. 2021. Noncoding RNA: an insight into chloroplast and mitochondrial gene expressions. *Life* 11: 49.

Castandet B, Hotto AM, Strickler SR, Stern DB. 2016. ChloroSeq, an optimized chloroplast RNA-Seq bioinformatics pipeline, reveals remodeling of the organellar transcriptome under heat stress. *G3 Genes|Genomes|Genetics* 6: 2817–2827.

Chang JH, Tong L. 2012. Mitochondrial poly(A) polymerase and polyadenylation. *Biochimica et Biophysica Acta* 1819: 992–997.

Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris KV, Morillon A et al. 2011. The reality of pervasive transcription. *PLoS Biology* 9: e1000625.

Crampton-Platt A, Yu DW, Zhou X, Vogler AP. 2016. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience* 5: s13742–016–0120–y.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.

Dietrich A, Wallet C, Iqbal RK, Gualberto JM, Lotfi F. 2015. Organellar non-coding RNAs: emerging regulation mechanisms. *Biochimie* 117: 48–62.

Fields PD, Waneka G, Naish M, Schatz MC, Henderson IR, Sloan DB. 2022. Complete sequence of a 641-kb insertion of mitochondrial DNA in the *Arabidopsis thaliana* nuclear genome. *Genome Biology and Evolution* 14: evac059.

Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, Brown S, Capodiferro MR, Al-Ajli FO, Ambrosini R et al. 2021. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biology* 22: 120.

Forni G, Puccio G, Bourguignon T, Evans T, Mantovani B, Rota-Stabelli O, Luchetti A. 2019. Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes. *Scientific Reports* 9: 14806.

Forsythe ES, Grover CE, Miller ER, Conover JL, Arick MA, Chavarro MC, Leal-Bertioli SC, Peterson DG, Sharbrough J, Wendel JF et al. 2022. Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels. *Proceedings of the National Academy of Sciences of the United States of America* 119: e2204187119.

Georg J, Honsel A, Voss B, Rennenberg H, Hess WR. 2010. A long antisense RNA in plant chloroplasts. *New Phytologist* 186: 615–22.

Ghareghani M, Porubsky D, Sanders AD, Meiers S, Eichler EE, Korbel JO, Marschall T. 2018. Strand-seq enables reliable separation of long reads by chromosome via expectation maximization. *Bioinformatics* 34: i115-i123.

Gray MW, Lukeš J, Archibald JM, Keeling PJ, Doolittle WF. 2010. Irremediable complexity? *Science* 330: 920–1.

Guilcher M, Liehrmann A, Seyman C, Blein T, Rigaill G, Castandet B, Delannoy E. 2021. Full Length Transcriptome Highlights the Coordination of Plastid Transcript Processing. *International Journal of Molecular Sciences* 22: 1297.

Hotto AM, Germain A, Stern DB. 2012. Plastid non-coding RNAs: emerging candidates for gene regulation. *Trends in Plant Science* 17: 737–744.

Hotto AM, Schmitz RJ, Fei Z, Ecker JR, Stern DB. 2011. Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the Arabidopsis transcriptome. *G3 Genes|Genomes|Genetics* 1: 559–570.

Kainth AS, Haddad GA, Hall JM, Ruthenburg AJ. 2023. Merging short and stranded long reads improves transcript assembly. *PLoS Computational Biology* 19: e1011576.

Kleine T, Maier UG, Leister D. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology* 60: 115–138.

Kobayashi G. 2023. Buried treasure in a public repository: mining mitochondrial genes of 32 annelid species from sequence reads deposited in the Sequence Read Archive (SRA). *PeerJ – Life and Environment* 11: e16446.

Kovaka S, Ou S, Jenike KM, Schatz MC. 2023. Approaching complete genomes, transcriptomes and epi-omes with accurate long-read sequencing. *Nature Methods* 20: 12–16.

Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood A-MJ, Haugen E, Bracken CP, Rackham O, Stamatoyannopoulos JA et al. 2011. The human mitochondrial transcriptome. *Cell* 146: 645–658.

Mukhopadhyay J, Hausner G. 2021. Organellar introns in fungi, algae, and plants. *Cells* 10: 2001.

Nachtigall PG, Grazziotin FG, Junqueira-de-Azevedo ILM. 2021. MITGARD: an automated pipeline for mitochondrial genome assembly in eukaryotic species using RNA-seq data. *Briefings in Bioinformatics* 22: bbaa429.

Palazzo AF, Gregory TR. 2014. The case for junk DNA. *PLoS Genetics* 10: e1004351.

Pozzi A, Dowling DK. 2019. The genomic origins of small mitochondrial RNAs: are they transcribed by the mitochondrial DNA or by mitochondrial pseudogenes within the nucleus (NUMTs)? *Genome Biology and Evolution* 11: 1883–1896.

Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, DePamphilis CW et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342: 1468–73.

Ruwe H, Gutmann B, Schmitz-Linneweber C, Small I, Kindgren P. 2019. The E domain of CRR2 participates in sequence-specific recognition of RNA in plastids. *New Phytologist* 222: 218–29.

Sanita Lima M, Rossi Paschoal A, Silva Domingues D, Smith DR. 2024a. Plant organelle genome pervasive transcription: functional noncoding transcriptomes? *Trends in Plant Science* S1360-1385. doi: https://doi.org/10.1016/j.tplants.2024.01.004

Sanita Lima M, Silva Domingues D, Rossi Paschoal A, Smith DR. 2024b. *Long-read RNA sequencing can probe organelle genome pervasive transcription*. Briefings in Functional Genomics elae026. https://doi.org/10.1093/bfgp/elae026

Sanita Lima M, Smith DR. 2017a. Pervasive, genome-wide transcription in the organelle genomes of diverse plastid-bearing protists. *G3 Genes|Genomes|Genetics* 7: 3789–3796.

Sanita Lima M, Smith DR. 2017b. Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. *Genome Biology and Evolution* 9: 2650–2657.

Sanita Lima M, Woods LC, Cartwright MW, Smith DR. 2016. The (in)complete organelle genome: exploring the use and non-use of available technologies for characterizing mitochondrial and plastid chromosomes. *Molecular Ecology Resources* 16: 1279–1286.

Schuster J, Ritchie ME, Gouil Q. 2023. Restrander: rapid orientation and artefact removal for long-read cDNA data. *NAR Genomics and Bioinformatics* 5. https://doi.org/10.1093/nargab/lqad108

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 Biology* 10: e1001241.

Smith DR. 2013. RNA-Seq data: a goldmine for organelle research. *Briefings in Functional Genomics* 12: 454–6.

Smith DR, Sanita Lima M. 2017. Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline. *Briefings in Bioinformatics* 18: 1012–1016.

Stern DB, Goldschmidt-Clermont M, Hanson MR. 2010. Chloroplast RNA metabolism. *Annual Review of Plant Biology* 61: 125–155.

Su HJ, Barkman TJ, Hao W, Jones SS, Naumann J, Skippington E, Wafula EK, Hu JM, Palmer JD, DePamphilis CW. 2019. Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proceedings of the National Academy of Sciences of the United States of America* 116: 934–43.

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M et al. 2023. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics* 24: 288.

van Bakel H, Nislow C, Blencowe BJ, Hughes TR. 2010. Most "dark matter" transcripts are associated with known genes. *PLoS Biology* 8: e1000371.

van Bakel H, Nislow C, Blencowe BJ, Hughes TR. 2011. Response to "The reality of pervasive transcription". *PLoS Biology* 9: e1001102.

Vera A, Sugiura M. 1994. A novel RNA gene in the tobacco plastid genome: its possible role in the maturation of 16S rRNA. *The EMBO Journal* 13: 2211–2217.

Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, Jin S, Wu Z. 2024. Plant organellar genomes: much done, much more to do. *Trends in Plant Science* doi: https://doi.org/10.1016/j.tplants.2023.12.014.

Wu Z, Stone JD, Storchova H, Sloan DB. 2015. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of angiosperm *Silene noctiflora*. *BMC Genomics* 16: 938.

Zhelyazkova P, Sharma CM, Forstner KU, Liere K, Vogel J, Borner T. 2012. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *The Plant Cell* 24: 123–136.

Zhou W, Armijos CE, Lee C, Lu R, Wang J, Ruhlman TA, Jansen RK, Jones AM, Jones CD. 2023. Plastid genome assembly using long-read data. *Molecular Ecology Resources* 23: 1442–1457.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sanita Lima, M., Silva Domingues, D., Rossi Paschoal, A. & Smith, D.R. (2024) Long-read RNA-Seq for the discovery of long noncoding and antisense RNAs in plant organelles. *Physiologia Plantarum*, 176(4), e14418. Available from: https://doi.org/10.1111/ppl.14418