Trends in Plant Science

CellPress

Forum

Pervasive transcription of plant organelle genomes: functional noncoding transcriptomes?

Matheus Sanita Lima [1,*], Alexandre Rossi Paschoal [2], Douglas Silva Domingues [3] and David Roy Smith [1,4,*,@]

Plant mitochondrial and plastid genomes typically show pervasive, genome-wide transcription. Little is known, however, about the utility of organelle noncoding RNAs, which often make up most of the transcriptome. Here, we suggest that long-read sequencing data combined with dedicated RNA databases could help identify putative functional organelle noncoding transcripts.

The noncoding organelle transcriptome

Plant mitochondrial and plastid DNAs (mtDNAs and ptDNAs) have been studied extensively for over 40 years, representing some of the first chromosomes to be completely sequenced [1]. There are now thousands of plant organelle DNAs deposited in the National Center for Biotechnology Information (NCBI), with more arriving each day. With so much data, one might argue that their novelty has long since vanished. Indeed, the remarkable diversity of plant organelle genomic architecture is so well described that it has almost become unremarkable. 'mtDNAs bigger than bacterial genomes…' Come on, we have heard it before [2]. 'Dozens of horizontally acquired genes…' No big deal [3]. 'Thousands of RNA editing sites…' Old news [4]. Nevertheless, we believe that these genomes still have a lot to teach us, especially about **noncoding RNAs (ncRNAs)** (see Glossary).

The organelle genomes of numerous land plants and algae show **pervasive, genome-wide transcription** [5]. This may not necessarily sound impressive, given that many mitochondrial and plastid transcripts are polycistronic. However, we should consider that the mitochondrial genomes of land plants, for instance, can span millions of nucleotides [2], of which only a small fraction are coding sequences. The same is true for certain green and red algal **plastomes** [1]. In other words, some plants are generating enormous quantities (>1 Mb) of noncoding organelle transcripts, mostly coming from intergenic regions [5]. However, the function (if any) of these transcripts within organelles is poorly understood. Plants, given their propensity for bloated organelle genomes, are particularly well suited for studying the nuances of this **noncoding transcriptome**, which could possibly reveal previously hidden regulatory processes, such as those associated with the **anterograde** and **retrograde** signaling pathways [6]. There is also the possibility that plants are just good at generating 'junk' organelle RNAs, as exemplified by 'the onion test' [7] (Box 1).

Much of our understanding of organelle ncRNA comes from the mapping of short-read RNA sequencing data (RNA-Seq) onto completely assembled plant organelle genomes [5]. A single Illumina RNA-Seq dataset from a land plant or alga, for instance, typically contains thousands to millions of mitochondrial- and plastid-derived reads, allowing for a quick and easy assessment of organelle transcriptomes. Whether the genome is big or small, these mapping analyses have consistently shown that plants transcribe nearly all their mtDNA and ptDNA. In fact, it is often possible to reconstruct entire organelle genomes from RNA-Seq data

Glossary

Anterograde: the cell signaling pathway from the nucleus to the organelles. Despite having their own genomes, most of the proteins (and other molecules) that operate within mitochondria and plastids are encoded in the nucleus.
D-foci: the discernable regions of the mammalian mitochondrial matrix where RNA degradation appears to happen. D-foci might be a subtype of **MRG**s.
Exapted: concept championed by Stephen Jay Gould and Elisabeth Vrba. An exapted trait is a trait that, over evolutionary time, serves a function co-opted only later in the fixation of the given trait. The most used example is feathers. Feathers might have been selected for temperature control but were later exapted for flight. Exapted traits can emerge either via neutral or adaptive processes.
Long-read sequencing: DNA and RNA sequencing technologies, including PacBio and Oxford Nanopore, that typically generate reads 10–100 kb in length. Often called 'third generation sequencing', long-read technologies have greatly improved their base-calling error rates in recent years and are quickly becoming the most used types of sequencing.
Mitochondrial RNA granules (MRGs): nonmature RNAs in mammalian mitochondria that exhibit a higher level (3D) organization. MRGs are membraneless but display a discernable localization within the mitochondrial matrix. Just like the nucleoids, MRGs (and D-foci) have distinctive proteomes.
Mitogenome: the genomes of mitochondria. Mitochondrial genomes are the most sequenced type of chromosomes.
Noncoding RNAs (ncRNAs): RNAs that do not encode proteins, transfer RNAs, or ribosomal RNAs. Some ncRNAs can have regulatory and/or structural roles, whereas other are likely functionless background 'noise' generated through the transcriptional process. Large quantities of ncRNAs can be generated in mitochondria and plastids through genome-wide, pervasive transcription.
Noncoding transcriptome: the full suite of ncRNAs generated by a given genome.
Nucleoids: organelle DNA complexed with proteins can be organized into highly compact speckles called nucleoids. In animals, for example, these are the sites for mtDNA replication, repair, and transcription.
Pervasive, genome-wide transcription: transcription that happens outside the traditional boundaries of a protein-, tRNA-, or rRNA-coding gene. Originally considered solely spurious (and highly undesired by the cell), pervasive transcription is now seen as a reliable source of functional ncRNAs.
Plastome: the genomes of plastids and all other plastid types (chloroplasts, chromoplasts, leucoplasts, etioplasts, gerontoplasts, and others). These plastid types are found in the diverse tissues of land plant species. Most non-photosynthetic species also have plastomes, but some parasitic plants and

non-photosynthetic free-living algae have plastids that lack plastomes.
Retrograde: the communication channel from organelles to the nucleus. The retrograde pathway is the most complex in plastid-bearing eukaryotes, as they have mitochondria and plastids communicating with the nucleus (and between the organelles).
Transcriptomic expansion: expansion in the amount of transcribed DNA for a given genome.

alone [5]. However, the short-read approach has its limitations, especially for investigating organelle ncRNAs. The repeat-rich nature of land plant **mitogenomes** and green algal plastomes means that a single Illumina read can map to hundreds of different locations in a genome. Moreover, having thousands of short overlapping segments of RNA-Seq can make it hard to decipher the start and end of individual transcripts and if they span entire intergenic regions or contain portions of coding RNA. Thankfully, the recent arrival of (third generation) **long-read sequencing** technologies [8] can alleviate many of these limitations, providing a fine-tuned picture of the organelle noncoding transcriptome.

Long-read high-throughput sequencing can help dissect organelle noncoding transcriptomes

The utility of long-read transcriptome sequencing is not just that it provides more information per read, but that the read itself can give insights into transcriptomic architecture because, unlike Illumina short reads, it can capture an intact molecule of RNA in its entirety (i.e., a complete transcript). In the case of plants, long-read RNA-Seq is particularly useful because, even in the largest organelle genomes, a long read will often span an entire intergenic region (and the abutting coding segments), leaving little doubt about mis-mappings due to repeats. Similarly, when multiple coding regions are captured on a single long read, one can sometimes glean information about putative antisense transcripts because plant organelle genes are often arranged in opposing transcriptional polarities. But is there enough publicly available long-read RNA-Seq data to start analyzing plant organelle transcriptomes?

As of November 2023, NCBI's Sequence Read Archive contains >3900 long-read transcriptomic datasets for streptophytes, spanning hundreds of species. About three-quarters of these are PacBio sequencing, with the remaining coming from Oxford Nanopore Technologies. Of course, this represents a tiny fraction (<1%) of the available RNA-Seq data for land plants and there are very few (<50) long-read datasets for green or red algae (although data are quickly accumulating for model algae, like *Chlamydomonas*). Still, the availability of nearly 4000 long-read datasets represents an excellent launch pad for exploring organelle transcriptomes and identifying hypothetical functional ncRNAs in plants; and they can easily be paired with short-read datasets to improve coverage and error correction.

Organelle ncRNAs of diverse sizes and types (e.g., microRNAs and long ncRNAs) have already been found in animals and plants [6,9]. Given this diversity, identifying genomic signatures that could serve as telltale signs of functional ncRNAs remains to be tested. Repeat-rich organelle genomes could also produce a myriad of almost indistinguishable functional transcripts that are lost amidst a veritable 'alphabet soup'. If anything, mitochondrial-derived peptides [10] are a testament to the creative power of organelle genomes. These peptides are derived from short open-reading frames hidden within the quintessentially streamlined human mitogenome. As more long-read data accrue, phylogenetically informed comparative studies will surely help scientists identify even more unique organelle transcripts, possibly distinct families of ncRNAs. By integrating these studies with synteny analyses, one could also uncover whether certain noncoding regions were **exapted** for generating regulatory and/or structural ncRNAs. As organelle genomes are organized into highly compact **nucleoids** [11], the possible function of long ncRNAs (produced by organelle and/or nuclear genomes) in shaping nucleoids should be investigated. Similarly, long ncRNAs are fundamental in the assemblage of diverse nuclear bodies [12], thus, they could exert similar functions in organelle systems. As ncRNAs might participate in the anterograde and retrograde signaling pathways [6,9], they could also have a role to play in plastid differentiation (Figure 1). Plant biologists studying organelle transcriptomes can look to the recent discoveries of mammalian mitochondrial and nuclear ncRNAs as a case in point (Figure 1). Long noncoding mitochondrial RNAs (ncmtRNAs) such as SncmtRNA and LIPCAR are good examples of where to begin [13].

One of the most important assets when searching for novel functions within nucleotide sequences is the availability of a large, well-curated reference database. This is especially true in the field of RNA biology, where databases like RNAcentral

Box 1. The onion test to the biological meaning of genome size variation

The onion test, developed by T. Ryan Gregory, posits that 'if most eukaryotic [nuclear] DNA is functional at the organism level, be it for gene regulation, protection against mutations, maintenance of chromosome structure, or any other such role, then why does an onion require five times more of it than a human?' [7]. The test could have been devised with any other species with a nuclear genome larger than human, but onions are a particularly good choice because among species their nuclear genomes range from 7 to 31.5 picograms, despite no perceptible changes in phenotypic complexity. The 'onion test' can easily be adapted to organelle genomes. For example, the mtDNA of the common onion is 19 times larger than the human mtDNA (316.4 kb versus 16.6 kb), even though it encodes only a handful more genes (accession number NC_030100.1). Simple logic would suggest that if the mitochondrion of human, which appears to be equally complex as that of onion, can get by with a fraction as much noncoding DNA, the excess mtDNA sequences in onion might well be junk.

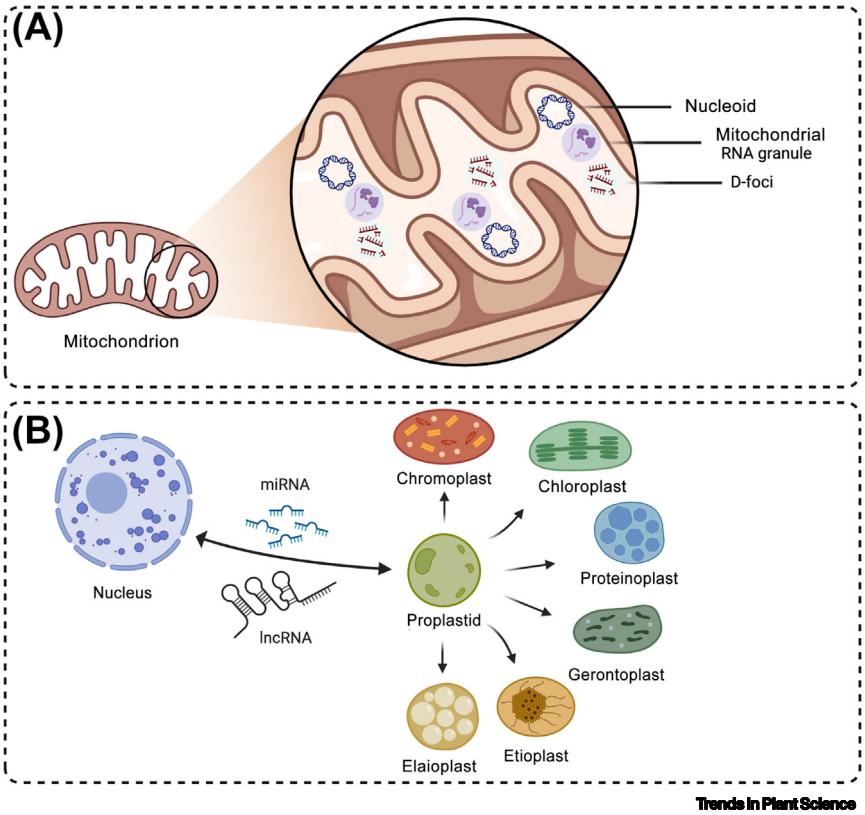

Figure 1. Possible ways through which noncoding RNAs (ncRNAs) (from the nucleus and from organelle genomes) can function. ncRNAs of various sizes are produced in both the nucleus and in organelles [6,9]. These ncRNAs might participate in the anterograde/retrograde pathways and in the high-level organization of organelle genomes. (A) Close-up of a mammalian mitochondrial matrix showing the nucleoids, **mitochondrial RNA granules (MRGs)**, and **D-foci** [11]. Nucleoids are the common configuration of mitogenomes and plastomes, but MRGs and D-foci have been found only in mammalian mitochondria so far. Long noncoding RNAs are known to help form nuclear bodies, such as Cajal bodies and paraspeckles [12]. Therefore, ncRNAs (particularly long noncoding RNAs) with structural function could be responsible for the assemblage of MRGs and D-foci in mitochondria and plastids of other lineages. (B) Plastid differentiation and fate in land plants is poorly understood, but it certainly relies on anterograde/retrograde signaling. We hypothesize that ncRNAs of various sizes could be contributing to the regulation of plastid differentiation. Note that these ncRNAs are expected to be produced by both genetic compartments (the nucleus and the plastids). Membraneless nuclear bodies (such as nuclear speckles [12]) are represented by the multiple circles inside the nucleus. Although not shown, the plastomes of the diverse plastid types are organized in nucleoids. RNA granules and D-foci have not been reported in plastids, but their presence (along with the presence of accessory ncRNAs) should be investigated. Abbreviations: miRNA, micro-RNA; lncRNA, long noncoding RNA. Created with BioRender.com.

(https://rnacentral.org/) have helped researchers identify a plethora of new ncRNA species. Unfortunately, there is not yet a database devoted to organelle RNAs. Moving forward, it will be crucial to develop such a database, focusing not just on transcripts encoding proteins, tRNAs, and rRNAs, but on those coming from intergenic, intronic, and antisense regions of organelle genomes as well. Admittedly, ncRNAs can have low sequence conservation, even among close relatives [14]. However, the large number of available plant organelle genomes and corresponding long-read datasets should provide enough raw data to start populating a database with putative ncRNA sequences. Other lineages should also be investigated, including those with compact genomes (e.g., animal mtDNA), some of which have already been shown to produce functional ncRNAs [6,9].

The focus need not be on function alone. Even if most organelle ncRNAs generated by plants and other species prove to be nonfunctional (or even slightly deleterious), they can still inform us about the processes driving **transcriptomic expansion**. Organelle DNAs are already a poster child for how nonadaptive forces can shape genome size and complexity. Their study has spurred various nonadaptive evolutionary theories, including the mutational hazard hypothesis and constructive neutral evolution [15]. Perhaps detailed analyses of the ncRNAs from mitochondria and plastids will have a similar impact on the field? Chopin famously said: 'Nothing is more odious than music without hidden meaning'. It remains to be determined if plant ncRNAs harbor hidden meanings. But generating greater interest in these sequences is music to our ears.

Declaration of interests

The authors declare no competing interests

[1]Department of Biology, Western University, London, Ontario, N6A 5B7, Canada
[2]Department of Computer Science, Bioinformatics and Pattern Recognition Group, Federal University of Technology – Paraná – UTFPR, Cornélio Procópio, PR, Brazil
[3]Department of Genetics, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, SP, Brazil
[4]https://www.arrogantgenome.com/

*Correspondence:
msanital@uwo.ca (M. Sanita Lima) and
dsmit242@uwo.ca (D.R. Smith).
@Twitter: @arrogantgenome (D.R. Smith).

https://doi.org/10.1016/j.tplants.2024.01.004

References

1. Smith, D.R. (2016) The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief. Funct. Genomics* 15, 47–54
2. Alverson, A.J. *et al.* (2011) Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* 23, 2499–2513
3. Garcia, L.E. *et al.* (2021) Horizontal gene transfers dominate the functional mitochondrial gene space of a holoparasitic plant. *New Phytol.* 229, 1701–1714
4. Smith, D.R. (2020) Unparalleled variation in RNA editing among *Selaginella* plastomes. *Plant Physiol.* 182, 12–14

5. Sanita Lima, M. and Smith, D.R. (2017) Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. *Genome Biol. Evol.* 9, 2650–2657
6. Dietrich, A. *et al.* (2015) Organellar non-coding RNAs: emerging regulation mechanisms. *Biochimie* 117, 48–62
7. Palazzo, A.F. and Gregory, T.R. (2014) The case for junk DNA. *PLoS Genet.* 10, e1004351
8. Kovaka, S. *et al.* (2023) Approaching complete genomes, transcriptomes and epi-omes with accurate long-read sequencing. *Nat. Methods* 20, 12–16
9. Anand, A. and Pandi, G. (2021) Noncoding RNA: an insight into chloroplast and mitochondrial gene expressions. *Life (Basel)* 11, 49
10. Merry, T.L. *et al.* (2020) Mitochondrial-derived peptides in energy metabolism. *Am. J. Physiol. Endocrinol. Metab.* 319, E659–E666
11. Pearce, S.F. *et al.* (2017) Regulation of mammalian mitochondrial gene expression: recent advances. *Trends Biochem. Sci.* 42, 625–639
12. Mattick, J.S. *et al.* (2023) Long non-coding RNAs: definitions, functions, challenges, and recommendations. *Nat. Rev. Mol. Cell Biol.* 24, 430–447
13. Ren, B. *et al.* (2023) Emerging functions of mitochondria-encoded noncoding RNAs. *Trends Genet.* 39, 125–139
14. Ulitsky, I. (2016) Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* 17, 601–614
15. Smith, D.R. (2016) The mutational hazard hypothesis of organelle genome evolution: ten years on. *Mol. Ecol.* 25, 3769–3775