Correspondence

The mitochondrial and chloroplast genomes of the green alga *Haematococcus* are made up of nearly identical repetitive sequences

Xi Zhang1, Nicholas Bauman1, Rob Brown2, Toby H. Richardson2, Srividya Akella1, Elizabeth Hann2, Robert Morey2, and David Roy Smith1,*

The chlamydomonadalean green alga *Haematococcus lacus* (strain UTEX 2505) has the largest chloroplast genome on record: 1352 kb with ~90% non-coding DNA [1,2]. But what of the mitochondrial genome? Here we present sequencing, assembly, and analysis of the mitogenome that shows that it, too, is extremely expanded. What’s more, the same repetitive elements have spread throughout the mitochondrial and chloroplast genomes, resulting in the situation whereby these two distinct organelle genomes are made up of nearly identical sequences.

At 124.6 kb, the *H. lacus* mtDNA is the biggest mitogenome yet observed from the Chlorophyta and among the largest of all plastid-bearing protists. Like the neighbouring ptDNA, the mtDNA is ~90% non-coding, a consequence of massive intergenic spacers (0.7–16.7 kb; average = 5.8 kb) rather than an abundance of introns, of which only three were identified (one of group I and two of group II affiliation). It is the sequences within these intergenic regions that make this mitochondrial genome so exceptional.

The non-coding mtDNA of *H. lacus* is almost entirely populated by repetitive elements, all based on a common motif (Figure 1A; see Supplemental Information). The repeats range in size from 30–700 nt, are GC-rich (average = 53.7%), and have a palindromic architecture, meaning that they can be folded into hairpin structures.

There are over 40 instances where long stretches (300–485 nt) of mtDNA match to the ptDNA. Indeed, a single subset of the non-coding mtDNA can map to hundreds of locations in the ptDNA, and vice versa (Figure 1D). If multiple hits within a query range are included, the mtDNA contains >40,000 regions of similarity to the ptDNA.

These findings are not the outcome of assembly artifacts. The *H. lacus* mitochondrial and chloroplast genome sequences were generated using solely long-read PacBio data (see Supplemental Information). Many of the mtDNA- and ptDNA-derived reads are anchored in coding regions, and some span entire intergenic spacers, reducing the potential for read misalignment or the co-assembly of chloroplast and mitochondrial repeats. As expected, assembly of the mitogenome and plastome gave circular-mapping molecules with high coverage (average read coverage for mtDNA was >500x) and the standard cohort of genes typically found in chlamydomonadalean algae. Moreover, no copies of the repeats were identified in any of the nuclear DNA-derived reads or contigs.

Organelle genomes frequently harbor repeats, especially palindromic ones [3], and in at least two cases a small number (<40) of short direct repeats have been found in both the mtDNA and ptDNA from a green alga [4,5]. However, to the best of our knowledge, this is the first example, from any species, of nearly identical repetitive elements proliferating throughout two different organelle genomes of the same cell.

How did this come to be? Angiosperm mitochondrial genomes often contain tracks of ptDNA-derived sequences, acquired through intercompartmental (or horizontal) DNA transfer [6]. However, *Haematococcus*, unlike angiosperms, has only a single cup-shaped chloroplast per cell [7], greatly reducing the potential for ptDNA-to-mtDNA transmission [6]. *Haematococcus* can have multiple mitochondria per cell [8], increasing the likelihood of mitochondrial-to-chloroplast DNA migration, which is a rare but not undocumented phenomenon in eukaryotes, and one that has recently been observed in green algae [5]. The *H. lacus* organelle repeats could have travelled from one genetic compartment to another by piggybacking on an intron. Organelle introns are mobile and known to move between mitochondria and chloroplasts, including in green algae, and inter-organellar intron transfer has been implicated in the spread of repeats [4]. All of the introns in the *H. lacus* mtDNA and ptDNA contain repeats. But no matter how these repeats first arrived, once they were there they had an aptitude for propagating themselves, leading to the most expanded pair of mitochondrial and chloroplast genomes uncovered to date. The elevated GC content of the repeats has also resulted in both the mtDNA and ptDNA having a GC composition of ~50%, a trait seen in only a handful of other species [9].

Much has been written about the spread of palindromic repeats throughout organelle DNA [10], but exactly how this occurs is poorly understood and likely differs within and among lineages. The organelle repeats in *H. lacus* show no obvious sequence identity to known transposable elements, but they do show similarity to palindromic repeats in the ptDNA of another closely related alga, *Stephanosphaera pluvialis* (GenBank accessions KT625299-409). Unfortunately, the mitogenome from this species has not been sequenced, and it remains unknown if the palindromes are also found in the mtDNA. If they
are, it might provide clues as to how identical repeat elements populated two distinct genetic compartments, as would population genomics data from other Haematococcus species. Whatever the case, the H. lacustris organelle DNAs provide a new take on organelle genome expansion and intercompartment genetic transfer, blurring the lines between mtDNA and ptDNA.

**SUPPLEMENTAL INFORMATION**

Supplemental information includes experimental procedures, one figure, one data file, author contributions, acknowledgements, and supplemental references and can be found with this article online at https://doi.org/10.1016/j.cub.2019.06.040.

**REFERENCES**


*Department of Biology, University of Western Ontario, London, ON, N6A-5B7, Canada.
*Synthetic Genomics Inc., La Jolla, California, 92037, USA.
*E-mail: dsmit242@uwo.ca

---

**Figure 1. Nucleotide sequence similarity between the Haematococcus lacustris mitochondrial and chloroplast genomes.**

(A) Nucleotide alignment and secondary structure of a 143-nt palindromic repeat from the mtDNA and ptDNA. This repeat element represents region 948,342–948,484 and 16,958–17,100 of the ptDNA and mtDNA, respectively. (B) Nucleotide dot-plot similarity matrix of the entire H. lacustris ptDNA (x-axis; 1352 kb) versus the entire mtDNA (y-axis; 124.6 kb). Black dots within the matrix represent sequence similarity between the two organelle DNAs (that is, the palindromic repeats). Plots were generated with JDotter using a plot size of 1000 bases/pixel and a sliding window size of 40 (see the Supplemental Information). (D) Regions of the mtDNA and ptDNA that are similar to one another based on BLAST results. Mitochondrial DNA was aligned to ptDNA (and vice versa) using blastn and the default settings at the National Center for Biotechnology Information (NCBI) with the exception that the number of matches to a query range was limited to 1 (see Supplemental Information).