

Short Communication

Not seeing the genomes for the DNA

David Roy Smith

Advance Access publication date 9 March 2012

Abstract

This Short Communication highlights the diversity of ‘secondary’ genome data (like mitochondrial and plastid genomes) that can be gleaned from next-generation sequencing projects, and encourages researchers to be mindful that these data are often as informative and useful as the ‘primary’ genome data.

Keywords: *next-generation sequencing; mitochondrial DNA; organelle genome; plastid DNA*

Outside the conference room, coffee and cookie in hand, she quietly tells me that they never planned to sequence the bacterial genomes that she had presented to the audience a few minutes earlier. ‘We were *trying* to get the genome of a green alga!’ she says, ‘but the culture was contaminated and when the sequencing data came back, out popped these unusual and interesting bacterial DNAs. So we dropped the alga and published the bacteria.’

Next-generation sequencing (NGS) technologies do not discriminate: if there is a nucleotide sequence in the sample—be it bacterial or green algal—they will find it, especially if it is present in multiple copies. This fact has helped my research on organelle genomes immensely. In the past it was difficult, time-consuming and expensive to sequence a mitochondrial or plastid genome, often requiring tedious methods to isolate organelles and enrich for organelle DNA. Now, with the advent of high-throughput sequencing, a single run of total DNA from a eukaryotic species on an NGS platform, such as Illumina or Roche 454, typically yields enough data to assemble organelle genomes with greater than 500-fold coverage [1–3]. This approach is fast and cheap, and it also means that the large numbers of scientists who are using NGS technologies to study eukaryotic nuclear genomes are generating huge quantities of organelle DNA sequence data in the process. The same principle holds for any other

‘secondary’ genomes, such as plasmids, that are present in the sample used for sequencing. However, the PI’s of genome projects often overlook organelle DNAs and secondary genomes as a whole even though these types of data can be as interesting and useful as the ‘primary’ genomic data.

Novel insights into mitochondrial and plastid genome evolution have come from assembling sequencing reads that were mined from eukaryotic nuclear genome projects [4–6]. In my own research on organelle genome architecture, I have taken advantage of the large amounts of raw sequencing data from NGS and Sanger platforms that are publically available in GenBank’s Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) and Trace Archive (www.ncbi.nlm.nih.gov/Traces/home/) to assemble the mitochondrial and plastid DNAs from diverse eukaryotic species [7]. It has been my own experience that whenever I have contacted the primary authors of these data, asking permission to use their sequences, they have been enthusiastic to collaborate; often they are surprised to hear that there are additional genomes within their data and that these genomes are worthy of analysis.

This strategy can help those from smaller institutions with limited resources collaborate with international research teams and be productive, all on a very limited budget. As more and more labs employ NGS technologies, I urge those who generate these

Corresponding author. David Roy Smith, Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada. E-mail: smithdr@dal.ca

David Roy Smith is a Izaak Walton Killam postdoctoral scholar in the Botany Department at the University of British Columbia, where he uses the mitochondrial and plastid genomes of protists to study the origins of genome architecture.

sequences to be mindful of the great variety of information that these data hold, and I encourage *all* researchers to actively explore these data, and make sure you do ‘see the genomes for the DNA’.

Key Points

- Pl's of eukaryotic nuclear genome projects often overlook organelle DNAs and ‘secondary’ genomes as a whole even though these types of data can be as interesting and useful as the ‘primary’ genomic data.
- Major insights into mitochondrial and plastid genome evolution have come from assembling sequencing reads that were mined from eukaryotic nuclear genome projects.
- I urge all researchers to be mindful of the great variety of information that NGS data hold and to actively explore these data, being cognizant of secondary genomes.

FUNDING

DRS is supported by postdoctoral fellowships from the Natural Sciences and Engineering Research Council of Canada and the Izaak Walton Killam Trusts.

References

1. Nock CJ, Waters DLE, Edwards MA, *et al.* Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J* 2011;**9**:328–33.
2. Wang W, Messing J. High-throughput sequencing of three *Lemnoideae* (duckweeds) chloroplast genomes from total DNA. *PLoS ONE* 2011;**6**:e24670.
3. Wolf PG, Der JP, Duffy AM, *et al.* The evolution of chloroplast genes and genomes in ferns. *Plant Mol Biol* 2010;**76**: 251–61.
4. Voigt O, Erpenbeck D, Wörheide G. A fragmented meta-zoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipapillata*. *BMC Genomics* 2008;**9**:350.
5. Hecht J, Grewe F, Knoop V. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early Tracheophytes. *Genome Biol Evol* 2011;**3**:344–58.
6. Oudot-Le Secq MP, Green BR. Complex repeat structures and novel features in the mitochondrial genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Gene* 2011;**476**:20–6.
7. Smith DR, Kayal E, Yanagihara AA, *et al.* First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented, linear architecture and insights into telomere evolution. *Genome Biol Evol* 2012;**4**:52–8.