MAY 10, 2023

Protocol Citation: Ramawatar Nagar 2023. a protocol for the detection mitochondrial heteroplasmy in a non model organism from whole genome sequencing data. **protocols.io** https://protocols.io/view/a-protocol-for-the-detection-mitochondrial-heterop-ctsbwnan

Protocol status: In development
We are still developing and optimizing this protocol

Created: May 05, 2023

Last Modified: May 10, 2023

PROTOCOL integer ID: 81443

Keywords: mitochondrial, heteroplasmy, Somatic hybridization

a protocol for the detection mitochondrial heteroplasmy in a non model organism from whole genome sequencing data

Ramawatar Nagar[1]

[1]ICAR-National Institute for Plant Biotechnology

Ramawatar Nagar
ICAR-National Institute for Plant Biotechnology

ABSTRACT

In the great majority of sexual eukaryotic organisms, mitochondrial inheritance is uniparental (usually maternal) which is tightly regulated to maintain homoplasmy in the progeny. However, many situations give rise to a cell with a mixture of two or more types of mitochondria, a state referred to as heteroplasmy. The situations where mitochondrial heteroplasmy can occur include:

- Somatic hybridization (exchange of cytoplasm).
- Biparental inheritance of mt-genomes.
- Paternal leakage of mitochondrial DNA.
- Random *de novo* mutations.

Somatic hybridization or anastomosis (hyphal fusion) has been reported in many fungal and plant species where cells of two diverse isolates fuse to produce offspring with a hybrid. In the event where two highly diverse mt-genomes come together in one cell leading to a high level of cytonuclear incompatibility, the cell selectively eliminates one of the mt-genomes eg for the *Rhizophagusirregularis* mushroom fungus. However, if there is no cytonuclear incompatibility, slightly different mt-genomes can co-exist within a single cell.
Detecting mitochondrial heteroplasmy in a non-model organism can be a challenging task as there may be limited genomic resources available. However, with the advancements in next-generation sequencing technologies, it is possible to sequence and assemble the mitochondrial genome even in non-model organisms. Here are some detailed steps for detecting mitochondrial heteroplasmy in a non-model species from whole-genome sequencing data

MATERIALS

Bioinformatics tools needed:

- a short read mapper such as BWA
- Seqtk for subsampling
- Freebayes for joint variant calling
- Trimmomatic or Cutadapt.
- MitoZ or MITObim
- GATK or SAMtools
- HaploGrep or MitoBamAnnotator

1 Mitochondrial genome assembly: Assemble the mitochondrial genome from the sequencing data using a reference genome or de novo assembly tools like MitoZ or MITObim. For non-model organisms, a reference genome may not be available. In such cases, you can use a related species as a reference or perform de novo assembly using the reads that map to the mitochondrial genome.

2 Mitochondrial genome annotation: Annotate the assembled mitochondrial genome using tools like MITOS or MFannot to identify the protein-coding genes, rRNA genes, tRNA genes, and other non-coding regions.

3 Quality control: Assess the quality of the sequencing data and remove any low-quality reads or adapters using tools like Trimmomatic or Cutadapt.

4 Read mapping: Map the reads back to the assembled mitochondrial genome using a read mapper such as BWA or Bowtie2.

5 Variant calling: Call variants (single nucleotide polymorphisms or insertions/deletions) using variant calling tools such as GATK or SAMtools. You can use the mitochondrial genome of a related species as a reference for calling variants. Be sure to use a high-quality variant calling pipeline to minimize false positives.

6 Heteroplasmy detection: Identify heteroplasmic variants by filtering out common variants found in a reference database or present in multiple individuals of the same species. To determine the level of heteroplasmy, you can use specialized software such as HaploGrep or MitoBamAnnotator that can calculate the frequency of each variant and estimate the heteroplasmy rate.

7 Validation: Validate the heteroplasmic variants detected through additional sequencing or PCR-based methods such as Sanger sequencing or high-throughput sequencing of PCR amplicons. It is important to ensure that the heteroplasmy is not due to sequencing errors or nuclear

mitochondrial DNA (NUMT) sequences.