First graft of *Talaromyces santanderensis'* mitochondria assembly

Juan José Picón Cossio & Juan Manuel Lozano Arias

Introduction

Talaromyces santanderensis is a new specie of fungus of which we do not know anything about its genetics. *T. santanderensis* has caught the attention of scientist because has a special condition; it is cadmium resistant. So, understand genetics of this fungus is valuable for the development of future biotechnology applications. In this work, given a few sequencing data collected using Oxford Nanopore MinION we are going to explore if it is possible to assembly a graft of *T. santanderensis'* mitochondrial genome.

Question

Is possible to assembly a graft of *T. santanderensis'* mitochondrial genome given a few reads? From 85.550 reads I will try to assembly a first graft of *T. santanderensis'* mitochondrial genome.

Approach & results

To achieve this goal, we first thought on a very useful tool known as GetOrganelle, but the primary function of this tool is to retrieve organelles, such as the mitochondria, from short reads like Ilumina reads, since we have Nanopore reads we thought that could not use it, however, exploring the tool we noticed that had another function to assembly the mitochondria given a genome assembly, so we decided to first assemble a graft of *T. santanderensis* genome, and in this way, subsequently we can use GetOrganelle. To assembly we used a tool called Flye, which is an assembler for long reads such as Nanopore reads.

```
flye --nano-hq talaromyces_reads.fastq -o talaromyces_ensam/ -t 32
```

The last command executed Flye and gave us an output with different files, between them was a file in gfa format (assembly_graph.gfa), which is an assembly graph format we are going to use as input for the tool GetOrganelle.

```
get_organelle_from_assembly.py -t 32 -g talaromyces_ensam/assembly_graph.gfa -F fungus_mt -o talaromyces_mt/
```

Once we ran GetOrganelle with the input (-g), database (-F) and output (-o) options, we received different files as output, among them were two main files, one in fasta format and other in gfa format. With the latter we can visualized the assembly (Image 1) and the former is the sequence itself to use in further analysis.

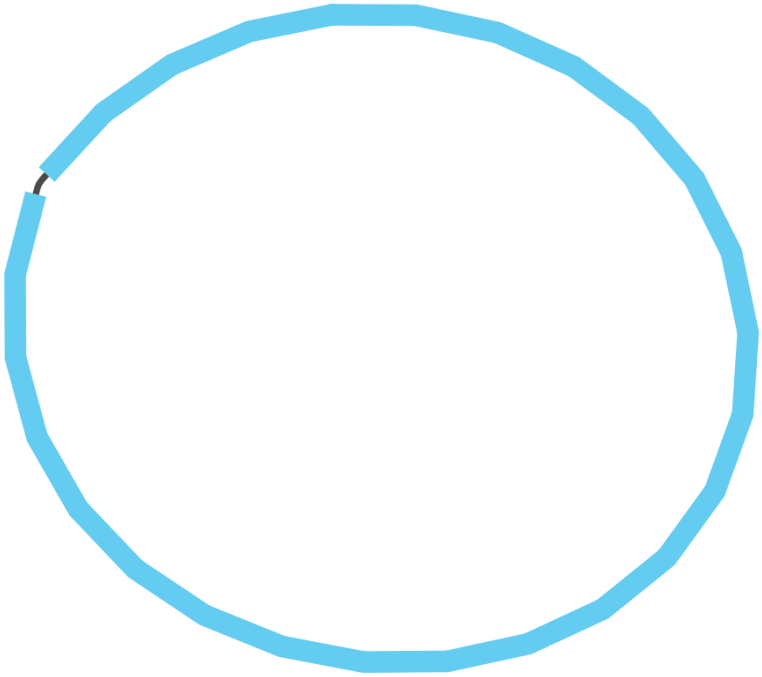

Image 1. Assembly graph.

Discussion

With the tool QUAST we calculated some statistics (Image 2). As you can see GetOrganelle assembled a circular genome, something expected, in one continuous contig of 31.379 pair bases, which result something very good, since *Talaromyces pinophilus* has approximately the same size of mitochondria (36.366 pb, https://www.ncbi.nlm.nih.gov/genome/?term=talaromyces+pinophilus)

```
All statistics are based on contigs of size >= 500 bp, unless otherwise noted
(e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs)
.

Assembly                    fungus_mt.complete.graph1.1.path_sequence
# contigs (>= 0 bp)         1
# contigs (>= 1000 bp)      1
# contigs (>= 5000 bp)      1
# contigs (>= 10000 bp)     1
# contigs (>= 25000 bp)     1
# contigs (>= 50000 bp)     0
Total length (>= 0 bp)      31379
Total length (>= 1000 bp)   31379
Total length (>= 5000 bp)   31379
Total length (>= 10000 bp)  31379
Total length (>= 25000 bp)  31379
Total length (>= 50000 bp)  0
# contigs                   1
Largest contig              31379
Total length                31379
GC (%)                      24.72
N50                         31379
N75                         31379
L50                         1
L75                         1
# N's per 100 kbp           0.00
```

Image 2. Assembly statistics.

Conclusions

Through this work we evidenced that a DNA extraction can have significant mitochondrial data and that, despite the little data collected, thanks to the size of the mitochondrial genome we can assemble, albeit unreliable, a draft mitochondrial genome. Additionally, we learned, albeit very basic, how to assemble a eukaryotic genome with nanopore data and how to assemble a mitochondrial genome.

References

1. Jian-Jun Jin*, Wen-Bin Yu*, Jun-Bo Yang, Yu Song, Claude W. dePamphilis, Ting-Shuang Yi, De-Zhu Li. **GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes.** *Genome Biology* **21**, 241 (2020). https://doi.org/10.1186/s13059-020-02154-5
2. Mikhail Kolmogorov, Jeffrey Yuan, Yu Lin and Pavel Pevzner, "Assembly of Long Error-Prone Reads Using Repeat Graphs", Nature Biotechnology, 2019 doi:10.1038/s41587-019-0072-8