



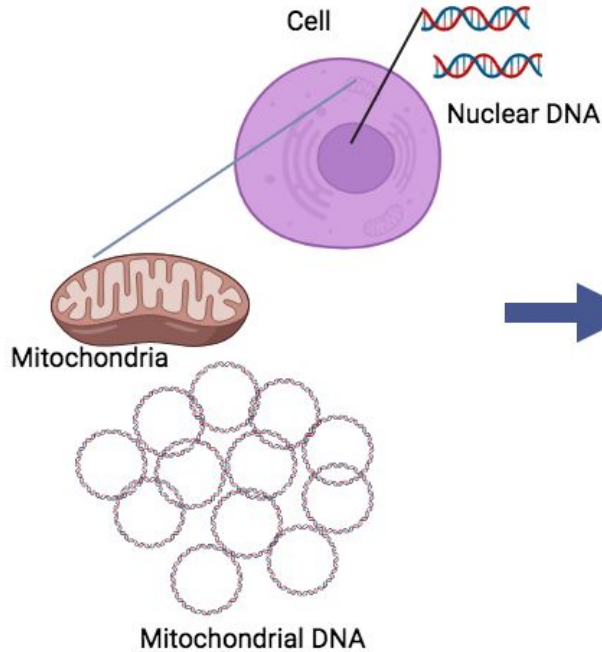
Mito-AssemblyVis: Mitochondrial Genome Assembly Assessment Visualization

Armaghan Sarvar, Cecilia Yang

**CPSC 547 Project Presentation
December 2022**



Mitochondrial Genome (Mitogenome)



DNA Barcodes for Environmental Survey

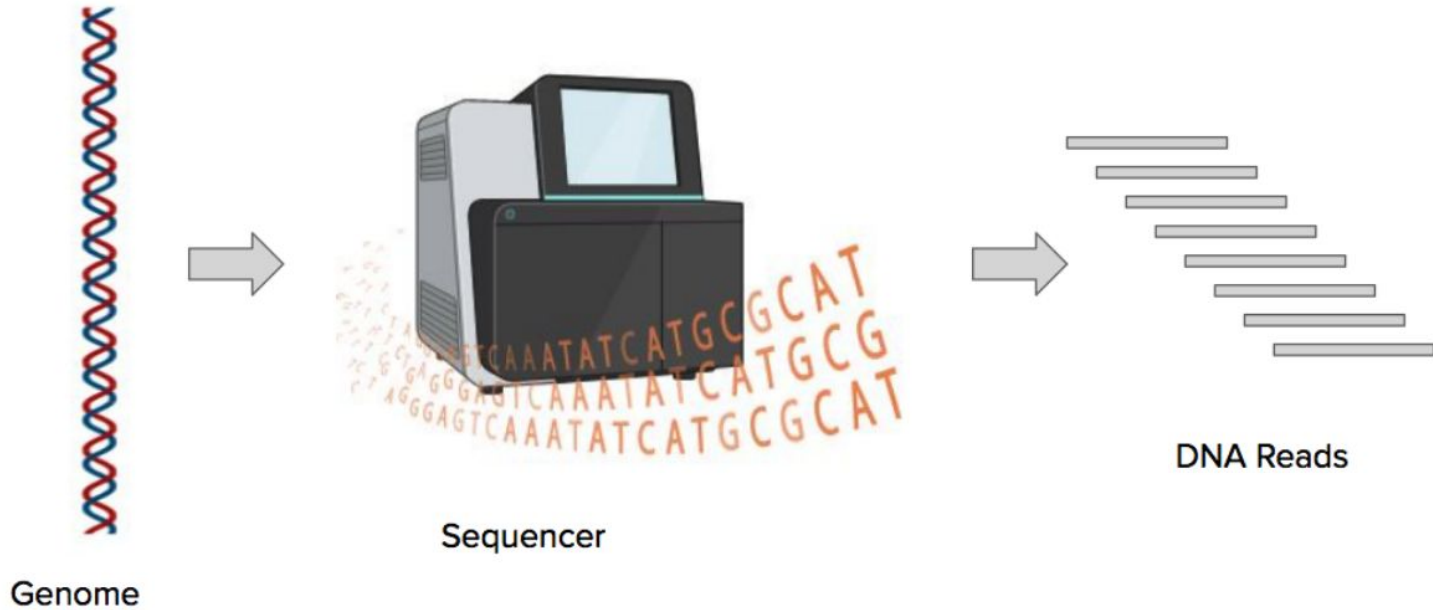
Used as DNA Finger Print in Forensics



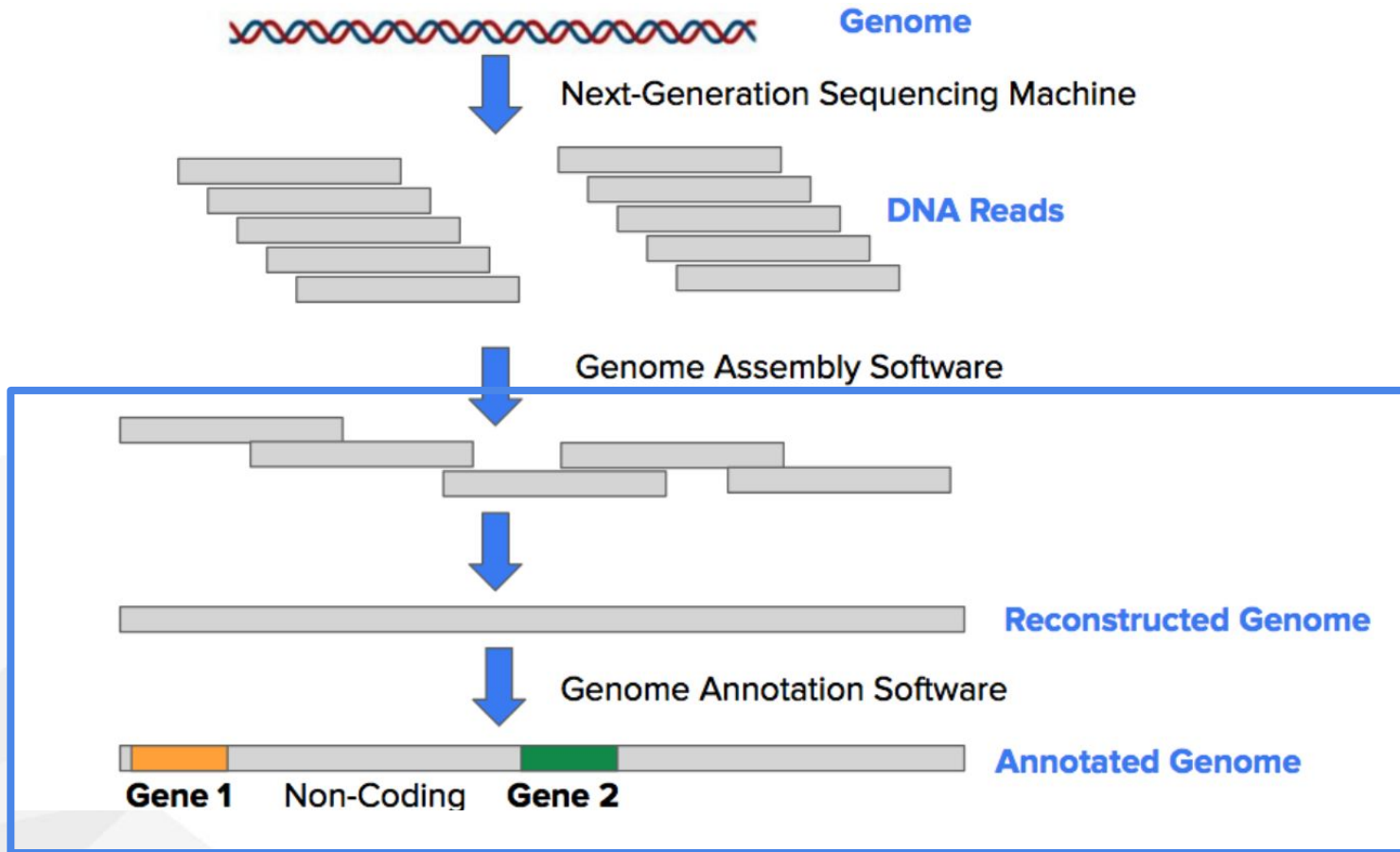
Anthropology Studies



Next-Generation DNA Sequencing



Genome Assembly and Annotation



Motivation



- There is currently a lack of tools designed specifically for mitogenome assembly visualizations
- Aimed to develop a visualization tool applicable to several assembly evaluation stages including fine tuning, annotation and genome similarity analysis



Task Abstraction

How good are the mitogenome assemblies?

Do these assembled sequences have desired lengths?

How identical are the assemblies compared to the reference sequences?

Were the optimal parameters chosen for the assembly pipeline softwares?

Are the expected set of genes found on the assembled genomes?

What are the gene orientations and the genome coverages?

Related Work



- Images created with Circos software, illustrating the relationships between different objects and regions

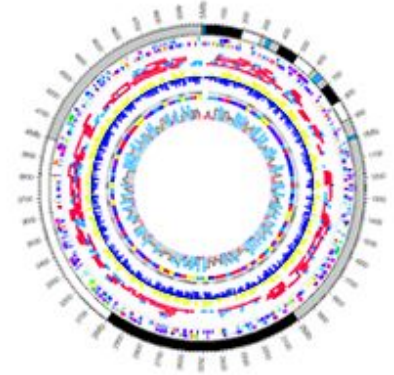
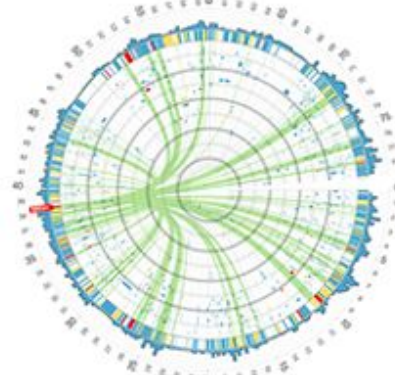
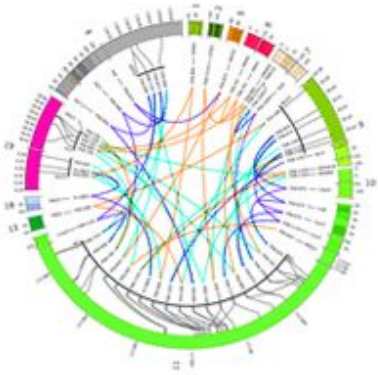
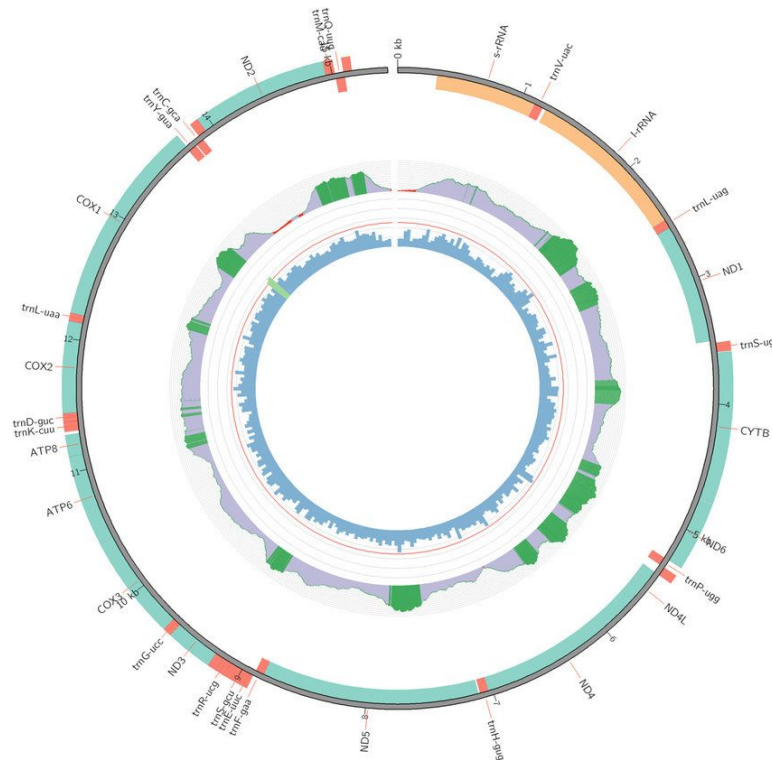


Image source: <http://circos.ca/>

Related Work



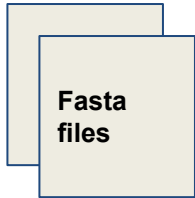
- Mitogenome Visualization using MitoZ



(Meng *et al.*, 2019)

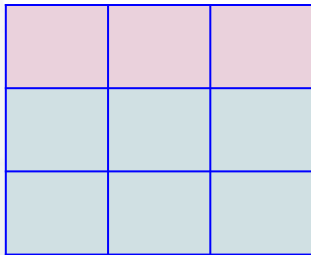
Data and Data Abstraction

Input Data



>Query sequence
ATGCCGT.....

>Reference sequence
ATGCGCT.....



Datasets containing quality metrics resulted from different assembly parameters, genome coverages

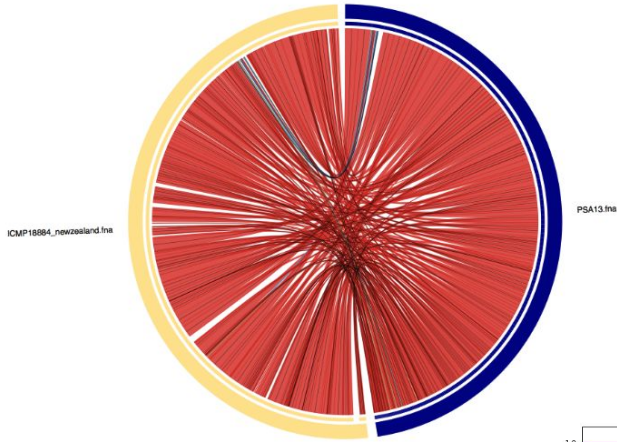


>Sequence 1; 1-79; +; Gene 1
CCGT.....

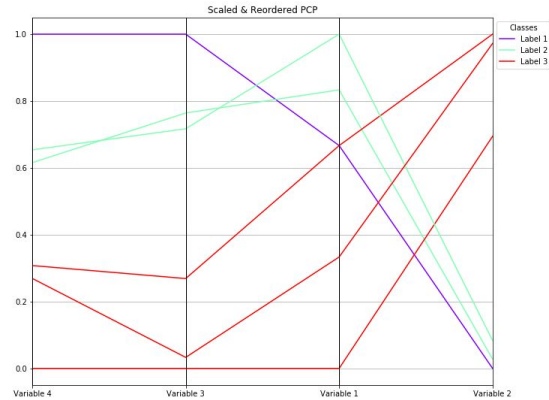
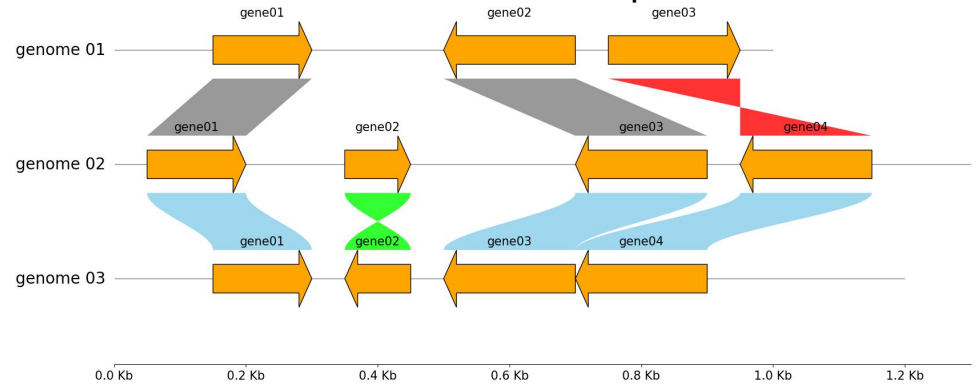
- Similarity between query and reference sequences (quantitative)
- Parameter(s) for assembler pipeline softwares and their corresponding quality metrics (quantitative)
- Gene location (continuous)
- Gene orientation (categorical)
- Genome coverage across the given genome (quantitative)

Solution

Circos Plot



Parallel Genome Annotation Comparison Plot



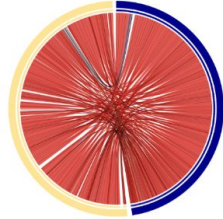
Parallel Coordinate Plot

Implementation of Mito-AssemblyVis



Flask

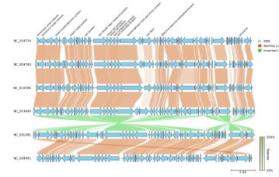
**Mito-AssemblyVis: Mitochondrial Genome
Assembly Assessment Visualization**



**GGisy-Based Circos
Plot**

HiPlot

**HiPlots for
Hyperparameter
Tuning**



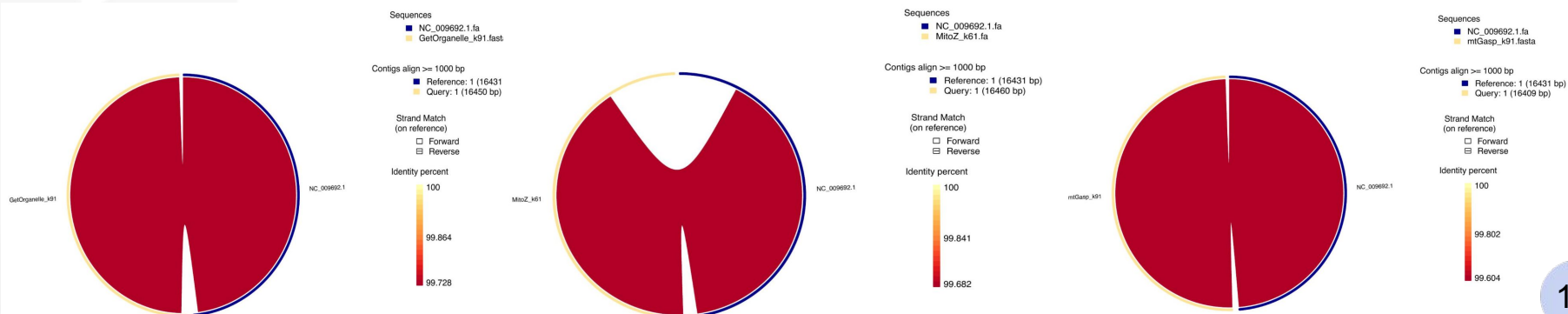
**PyGenomeViz
Annotation
Visualization**



<https://github.com/bcgsc/Mito-AssemblyVis>

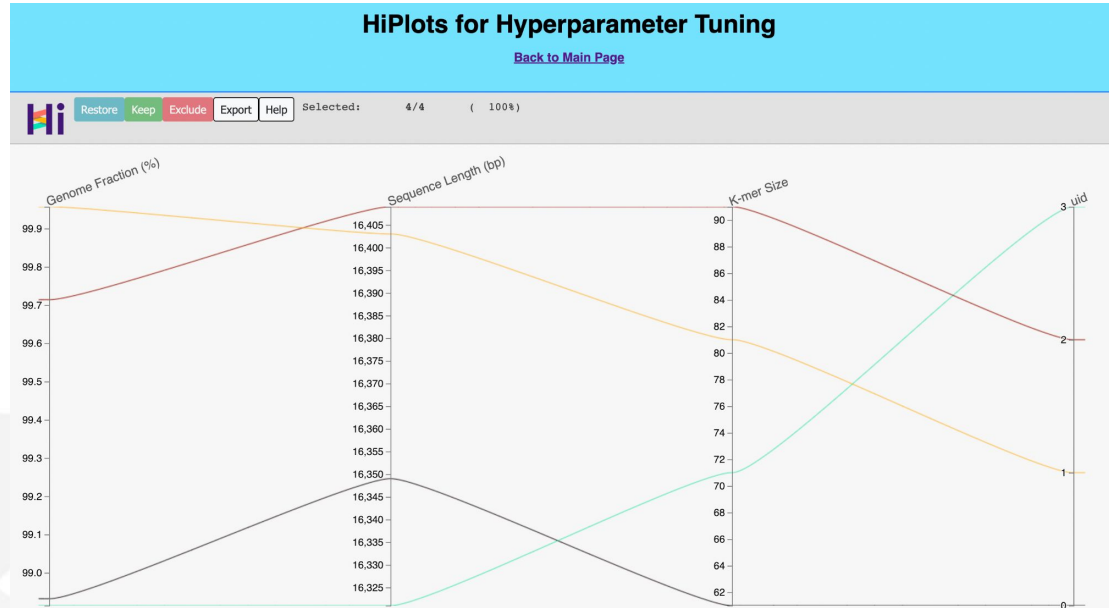
Visualization 1: GGisy-Based Circos Plot

- ? How identical is an **assembly** compared to a **reference** genomic sequence?
- ? How **complete** are the results of different assembly pipelines?
- ☑ Input: a set of contigs relative to the reference genome
 - ➡ Output: Circos-based genome assembly consistency plot
- ☑ Quick qualitative view of the misassemblies and gaps in a genome assembly



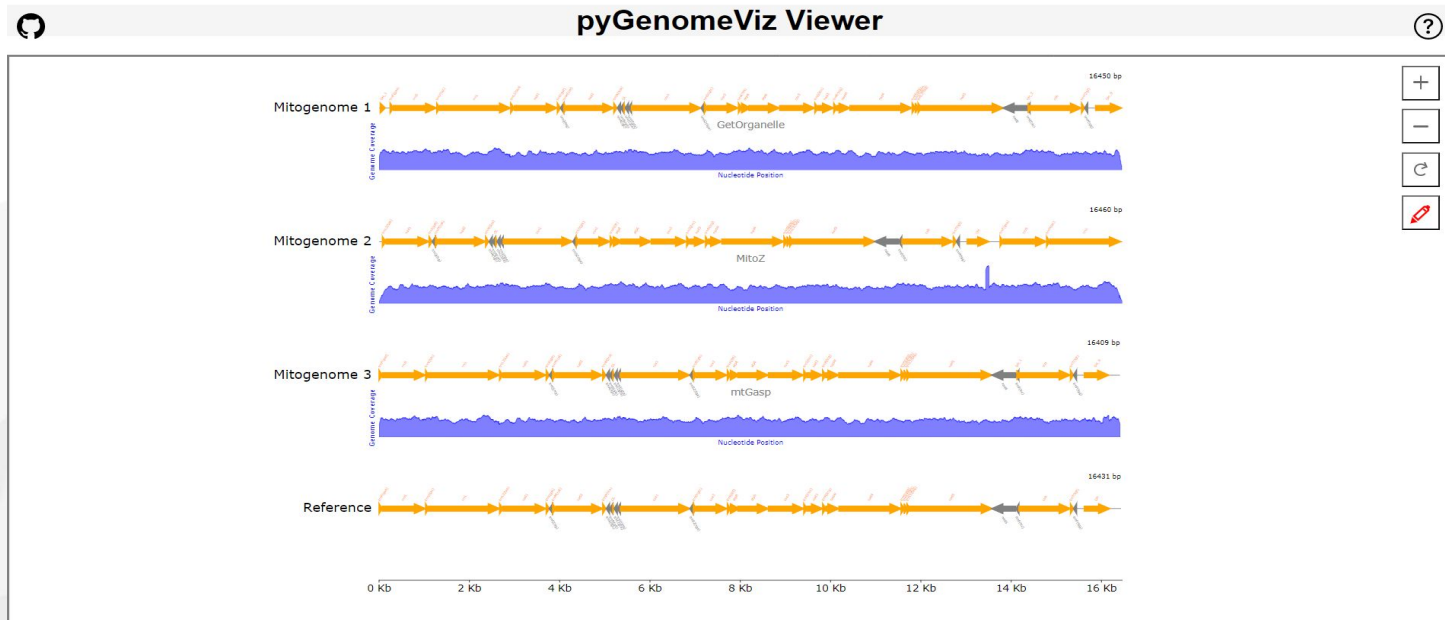
Visualization 2: HiPlot Parameter Tuning

- ? What is the best assembly parameter?
- ✓ Parallel coordinate plot facilitates the evaluation of multiple assembly quality metrics



Visualization 3: PyGenomeViz Parallel Annotation Comparison

- ❓ Are the expected set of genes found on the assembled genomes?
- ✅ Parallel annotation plot allows users to compare the gene annotation results of different assemblies



The background features a central white circle containing the text 'Application Demo'. This circle is surrounded by a ring of overlapping, semi-transparent geometric shapes in various colors including blue, green, purple, and orange. Additional clusters of these shapes are visible in the top-left and bottom-right corners of the slide.

Application Demo

Conclusions

- ✓ We propose a multi-purpose visualization interface specifically for mitogenome assembly assessment to facilitate the **assembly parameter tunings, sequence comparisons** and **annotation analysis**
- ✓ Users are able to generate customized visualizations by providing their generated assembly data files to Mito-AssemblyVis
- ✓ Interactive features of HiPlot and pyGenomeViz allow users to highlight the important attributes and navigate through the data in details

Limitations and Future Work



- A web-based application may not be ideal when working with larger genomic datasets. For such cases, a desktop application may be preferred.
- Host the application on the internet instead of the local machine
- The proposed tool has only been tested on one set of assembly data generated from a single whole genome sequencing (WGS) library due to time constraints.
- In the future, the generalizability of the tool will be investigated by visualizing different sets of mitogenome assembly results from other WGS libraries.
- Improvements in each module:
 - GGisy-based plot can be improved to also show more details such as small misassemblies, possibly mediated by repeats.

References

Facebook Research. *HiPlot 0.1.33*. Retrieved December 12, 2022, from <https://facebookresearch.github.io/hiplot/>

Flask Documentation (2.2.x). Retrieved December 12, 2022, from <https://flask.palletsprojects.com/en/2.2.x/>

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9), 1639–1645. <https://doi.org/10.1101/gr.092759.109>

Meng, G., Li, Y., Yang, C., & Liu, S. (2019). Mitoz: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, 47(11). <https://doi.org/10.1093/nar/gkz173>

Moshi4. *Pygenomeviz*. Retrieved December 12, 2022, from <https://pypi.org/project/pygenomeviz/>

Sanrrone. *GGISY: Genome-genome circle synteny*. GitHub. Retrieved December 12, 2022, from <https://github.com/Sanrrone/GGisy>



Thank you