1. Title : Imputation of Ancient Mitogenome Data Using Computational Models
2. Names :

* Yousef Abugosh : 315177444, [yousef.abu-gosh@mail.huji.ac.il](mailto:yousef.abu-gosh@mail.huji.ac.il), yousefak
* Areen Mansour : 212900211, [areen.mansour@mail.huji.ac.il](mailto:areen.mansour@mail.huji.ac.il), areen0507
* Nosiba Othman : 212378111, [nosiba.otman@mail.huji.ac.il](mailto:nosiba.otman@mail.huji.ac.il), nosiba
* Adan Mhameed : 214310666, [adan.mhameed2@mail.huji.ac.il](mailto:adan.mhameed2@mail.huji.ac.il), adan.mhameed2

1. Background: The research addresses the challenge of reconstructing complete mitochondrial genomes from ancient DNA samples with low-depth sequencing data. The problem arises due to degradation and insufficient sequencing depth, common in archaeological samples. The imputation method described in the referenced article provides a computational solution for filling missing nucleotide sequences.

Reference article:

- Title: Imputation approach for deducing a complete mitogenome sequence from low-depth-coverage next-generation sequencing data: application to ancient remains from the Moon Pyramid, Mexico

- Authors: Fuzuki Mizuno et al.

- Published: February 16, 2017, in Journal of Human Genetics.

- Summary: The study presents an imputation method using population-specific panels to reconstruct degraded mitochondrial sequences.

1. Research Question and Objectives:

* Question: How can computational imputation methods be optimized to enhance the accuracy of ancient DNA sequence reconstruction?
* Objective:
  + To propose improvements to computational models for handling degraded ancient DNA .
  + To implement and optimize a k-nearest neighbor-based imputation method for ancient mitogenome reconstruction.
  + To evaluate the performance of population-specific and global reference panels in filling missing nucleotide sequences.
  + To analyze the accuracy and sensitivity of imputation using simulated datasets with varying levels of missing data.

1. Data Description:

The dataset includes:

-Low-depth coverage sequencing data from a 1500-year-old individual excavated at the Moon Pyramid, Teotihuacan, Mexico.

- Reference panels comprising worldwide and indigenous population mitochondrial genomes.

- Data format: FASTQ and aligned BAM files.

Data source:

- Published datasets and supplementary materials from the referenced article.

1. Hypothesis:

Using population-specific reference panels improves imputation accuracy for ancient DNA sequences compared to generic panels, offering insights into population ancestry and genomic integrity.

1. Computational and Statistical Methods:

Imputation Algorithm: K-nearest neighbor-based imputation.

Reference Panels: Evaluation of worldwide and population-specific panels.

Software Tools: MAFFT for sequence alignment, SAMtools for SNP identification.

Statistical Tests: Accuracy and sensitivity analysis of imputed sequences using simulated data with varying levels of missingness.

1. Execution Plan:

Dataset Preparation: Align raw sequencing data to the mitochondrial reference genome.

Imputation: Implement the k-nearest neighbor algorithm using various reference panels.

Validation: Compare imputed sequences with experimental results and assess accuracy.

Analysis: Evaluate the impact of different panels on imputation performance.

Team Coordination: Assign team members specific tasks such as alignment, imputation, and statistical analysis.

1. References:

* Mizuno, F., Kumagai, M., et al. (2017). Imputation approach for deducing a complete mitogenome sequence... Journal of Human Genetics, 62, 631–635. DOI:10.1038/jhg.2017.14.
* <https://github.com/omics-tools/mitoimp/tree/master>.

1. We are a heterogeneous group of males and females, one neuro scientist and computer science Engineer, three computer science Engineers.