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

* Yousef Abugosh : 315177444
* Areen Mansour : 212900211
* Nosiba Othman : 212378111
* Adan Mhameed : 214310666

1. 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.
2. To fill in \*Exercise 4\*, I need to summarize and adapt the provided details into the hackathon proposal structure outlined in the document. Here's a draft plan:
3. 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.

1. Data source:

- Published datasets and supplementary materials from the referenced article.

1. ### 6. 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. ### 7. Computational and Statistical Methods:
2. - \*Imputation Algorithm:\* K-nearest neighbor-based imputation.
3. - \*Reference Panels:\* Evaluation of worldwide and population-specific panels.
4. - \*Software Tools:\* MAFFT for sequence alignment, SAMtools for SNP identification.
5. - \*Statistical Tests:\* Accuracy and sensitivity analysis of imputed sequences using simulated data with varying levels of missingness.
6. ### 8. Execution Plan:
7. 1. \*Dataset Preparation:\* Align raw sequencing data to the mitochondrial reference genome.
8. 2. \*Imputation:\* Implement the k-nearest neighbor algorithm using various reference panels.
9. 3. \*Validation:\* Compare imputed sequences with experimental results and assess accuracy.
10. 4. \*Analysis:\* Evaluate the impact of different panels on imputation performance.
11. 5. \*Team Coordination:\* Assign team members specific tasks such as alignment, imputation, and statistical analysis.
12. ### 9. References:
13. - 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.
14. ---
15. Let me know if you'd like me to refine or adjust any section further!