Name of team: H.J.L.

List of team members:

- Heather Han (hhan16@jhu.edu)
- Adrian Johnston (ajohn277@jhu.edu)
- Zichen Liu (zliu98@jhu.edu)

Title: Zeros and Ones to ATGC's

Description:
We will evaluate and improve the methods of storing binary data as DNA sequences. This includes evaluating the differences in compression order on performance. For instance:

Compress data → encode → store as DNA
Or
Data → encode → compress → store

Specifically, our dataset will be from a video file to determine if the video quality is maintained after encoding and decoding using our compression methods.

Papers:

1. DNA Fountain enables a robust and efficient storage architecture, Erlich and Zielinski (2017). http://science.sciencemag.org/content/355/6328/950
2. Boiler: lossy compression of RNA-seq alignments using coverage vectors (2016). https://academic.oup.com/nar/article/44/16/e133/2460171
3. Large-scale compression of genomic sequence databases with the Burrows–Wheeler transform (2012). https://academic.oup.com/bioinformatics/article/28/11/1415/266914

Datasets:
https://commons.wikimedia.org/wiki/File:DNA_packing.theora.ogv
(maybe more videos if time permits)

Access to MARCC:
Not sure how this works, but would we be able to have access *just in case*? We are not sure if we will need it, but we may.