Emily Price
Capstone Abstract
SAMURS

SAMURS Capstone Abstract

Duplicated genes and other nuclear genomic features typically diverge over time in sequence composition due to random mutations and other sources of genetic change. Surprisingly, mitochondrial genomes with multiple genes or control regions have a tendency or ability to consistently preserve near or complete identicality to each other (despite naturally occurring mutations), indicating either mutation reversal or other genetic mechanisms that allow them to be consistently identical, an unusual property that may hold key information on possible mutation reversal within these duplicated regions. To better understand this process across eukaryotes, all publically available mitochondrial genomes (12,370 total at time of study) were downloaded from the National Center for Biotechnology Information (NCBI) online database. These files were used in conjunction with a custom Python script that was created in order to parse the genome files and identify genes and possible control regions, as well as their frequency; these frequencies ranged up to two or more genes or control regions. Pairwise local sequence alignment is currently in progress (specifically, the EMBOSS water tool) in combination with another custom Python script to automate identification of duplicated regions for all genes and possible control regions for each available species (approximately 15 million comparisons in total), as well as to find the extent in which these genes and control regions are identical to each other in any given mitochondrial genome. Work is still ongoing and will be completed by spring 2023.