

Investigations of Duplications and Possible Mutation Reversal in Mitochondrial Genome Control Regions

Eric Edsinger and Emily Price

BACKGROUND AND INTRODUCTION

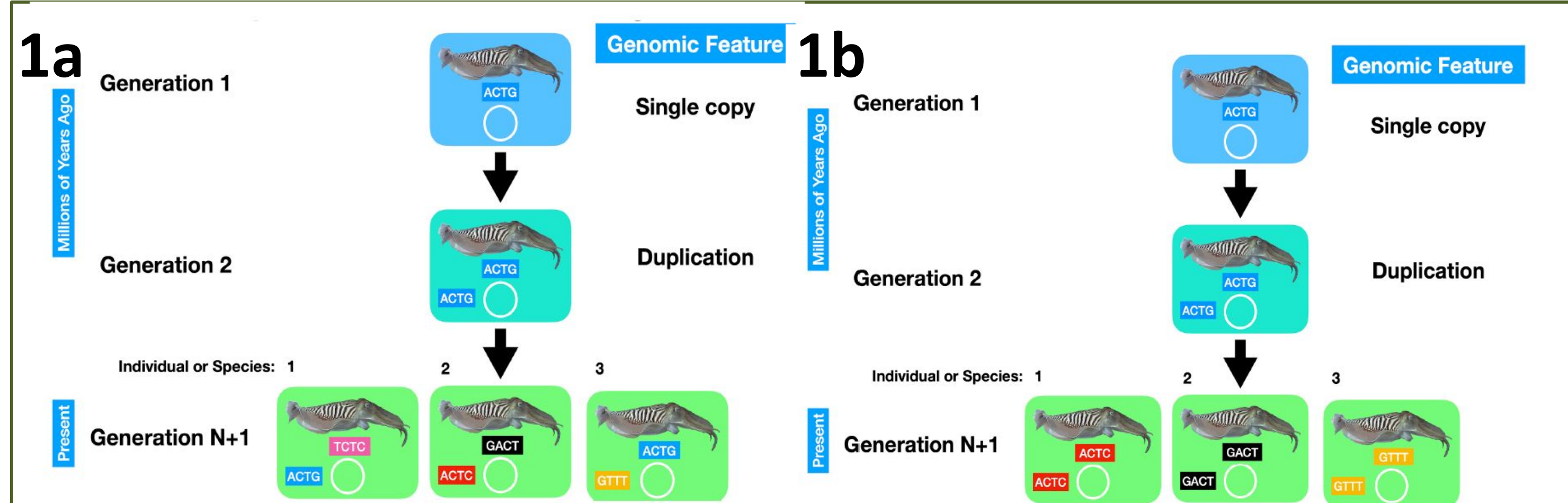
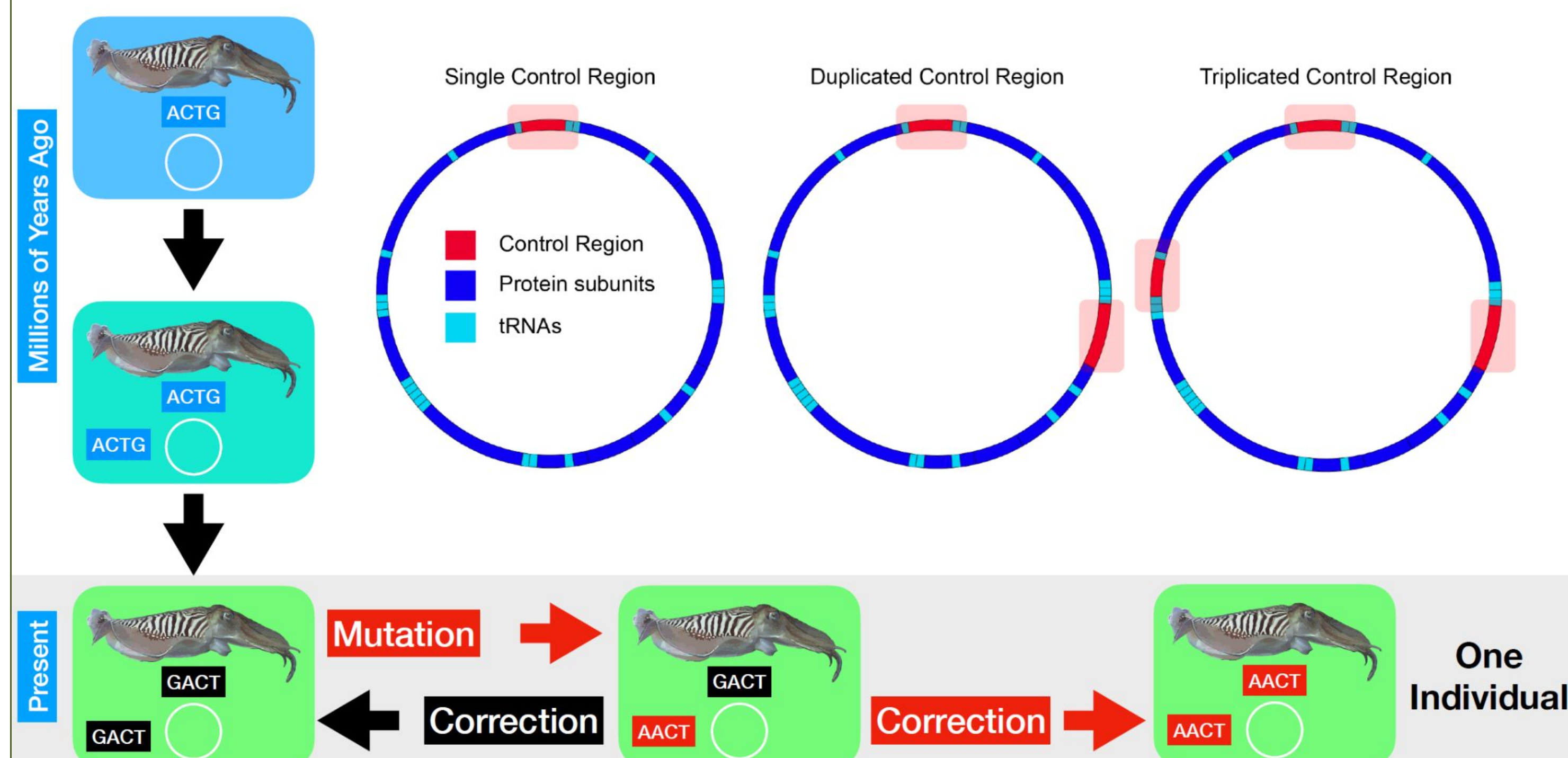


Figure 1. Illustrations showing effects of gene mutation on normal genomes (1a) compared to control region mutation reversal on mitochondrial genomes with duplicated or triplicated control regions (1b). Image credit: Kaila Rosales

In most genomes, genes and other features tend to occasionally acquire mutations, resulting in change over time. Though, in mitochondrial genomes with multiple control regions, these control regions and genes maintain identity to each other (Figure 1).

There are two possible genetic mechanisms for how these control regions preserve similarity: mutation reversal and/or gene correction of other control regions to match the newly mutated region (Figure 2).

Figure 2. Illustration showing different possible mechanisms for how multiple control regions stay identical to each other. Image credit: Kaila Rosales



OBJECTIVES

To better understand the genetic mechanisms and happenings behind this phenomenon.

To investigate whether mutation reversal, gene correction, or both, is involved.

RESULTS

The project is in progress and still ongoing, meaning results and conclusions are not finalized at this point.

This project is estimated to be completed by Spring 2023.

METHODS AND MATERIALS

I wrote python scripts that analyze and parse genome files, as well as find duplicated genes and control regions.

Functions of scripts:

- take in 2 input files per species for both gene code and location
- identified possible control regions between genes, separated them into isolated files, and flagged species with more than one possible control region (Figure 3a).
- used these possible control regions, along with the genome features, and conducted pairwise local sequence alignment to determine similarity (Figure 3b).
- performed above objectives for 12,370 total eukaryotic species

Figure 3. Two output examples of script function. Image credit: Emily Price

```
SUCCESS!
Run complete. 12370 total species analyzed. Control region test size: 500

1483 species have multiple flagged control regions.
8998 species have a single flagged control region.
1886 species have no flagged control regions.
0 species have no available genes.
Species skipped: 3

#-----
# Aligned_sequences: 2
# 1: Bagrus_docmak-mitochondrion-NC_061386.1-feature_MN607_mgt11_start-4505_stop-4577_(2022-11-21)
# 2: Bagrus_docmak-mitochondrion-NC_061386.1-feature_MN607_mgt04_start-1378_stop-1447_(2022-11-21)
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 79
# Identity: 47/79 (59.5%)
# Similarity: 47/79 (59.5%)
# Gaps: 26/79 (32.9%)
# Score: 131.5
#-----
3a
```

ACKNOWLEDGEMENTS & SPECIAL THANKS

- Friday Harbor Labs and the 2022 FHL Biodiversity Genomics Apprenticeship
- Kaila Rosales