

# Project --- Comparative analysis of bacterial genomes

By Ginny Wu, Yifan Yang

*Input:* a collection of mitochondrial or chloroplast genomes from various bacterial species.

*Output:* Phylogeny of bacterial strains based on RNA alignment of mitochondrial or chloroplast sequences.

## Dependencies

- Biopython
- Tkinter
- HMMER
- rRNA\_hmm\_fs

## Usage

1. Download repository from Github
2. With a set of assembled mitochondria/chloroplast genome reads either:
  - a. input to NCBI Blastn and download Fasta files for rRNA of top matches
  - b. Input sequences into rRNA prediction program rRNA\_hmm\_fs
3. Run python script rRNA\_phylo.py (> python rRNA\_phylo.py)
4. Window will pop up to select rRNA one or more of the rRNA Fasta sequences
  - a. Script runs muscle alignment and calculates phylogeny based on rRNA sequences
5. Output: out.txt
  - a. Phylogeny tree and distance matrix of selected rRNA sequences

## Significance:

- [Chloroplast genomes: diversity, evolution, and applications in genetic engineering](#) (2016)
  - a. Advancement of genetic engineering has sparked research into tailoring organelle genomes. Our comparative analysis program can be used to assist in similar research.
  - b. There is also research into the endosymbiotic theory of mitochondria and chloroplast. Our comparative analysis program can be used to find similarities in

the genomes of various species with higher accuracy, data which can be used to induce gene transfer or evolutionary relationship.

- [Complete chloroplast genome sequence and comparative analysis of loblolly pine \(\*Pinus taeda\* L.\) with related species](#) (2018)
  - a. Extract and analyze the whole chloroplast genome.
  - b. There is also a determination of phylogenetic relationships of those bacteria in series. Our final goal can follow this, convergent or divergent/high degree of sequence similarity.

## Citations

- [HMMER](#)
  - HMMER 3.2.1 (March 2019); <http://hmmer.org/> # Copyright (C) 2011 Howard Hughes Medical Institute. # Freely distributed under the GNU General Public License (GPLv3).
- [rRNA\\_hmm\\_fs](#)
  - Ying Huang, Paul Gilna and Weizhong Li. " Identification of ribosomal RNA genes in metagenomic fragments". [Bioinformatics \(2009\) 25:1338-1340](#)
- [NCBI Blast](#)