



Mitochondrial DNA Extraction

And Analysis From High Throughput Sequencing Data

Alisha Mechtley

Department of Scientific Computing

Florida State University

Evolution 2016



Properties of an Ideal Sequence for Phylogenetic study

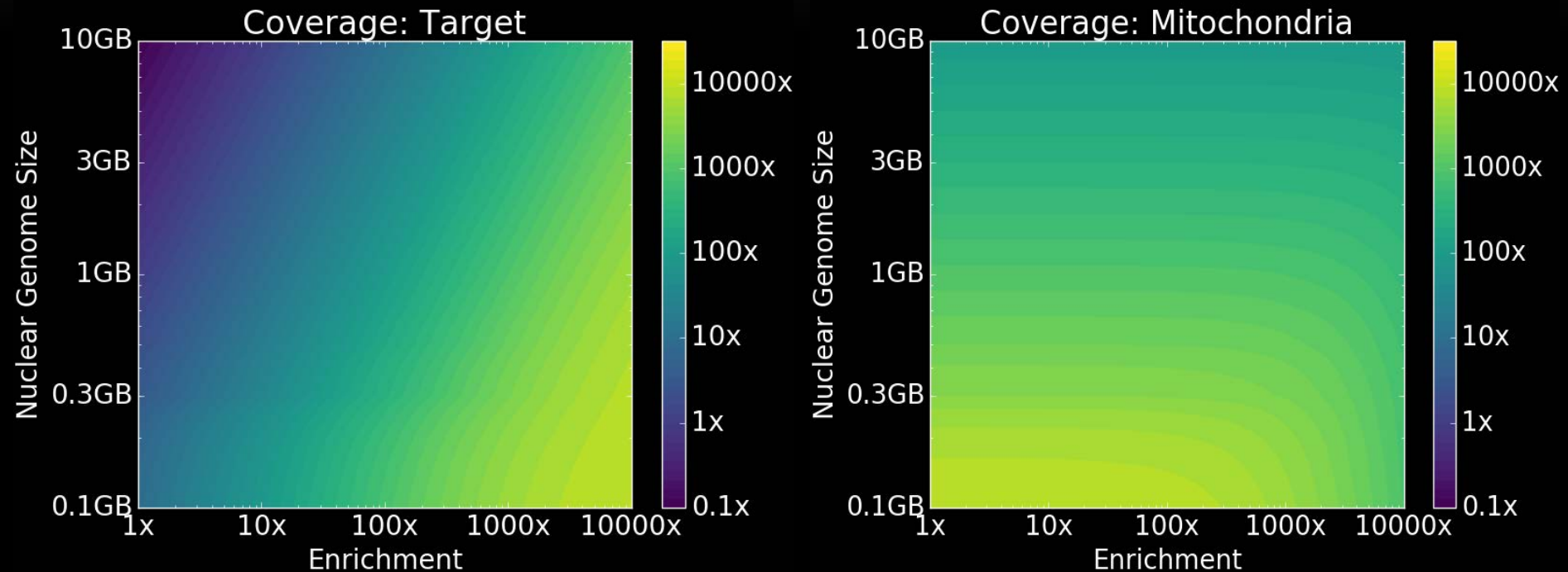
- a) Lack complicated structure
- b) Straightforward mode of genetic transmission
- c) High mutation rate
- d) Easy to isolate
- e) Distinguishable yet commonly distributed among a wide variety of organisms
- f) The major limitation is that it is effectively one gene.



Enrichment

- The cost of sequencing has decreased dramatically, and there is more data than necessary to answer many phylogenetic questions.
- As data quantity increases, so do the potential sources of phylogenetic error.
- Use a subset of data: Enrichment (e.g. Anchored Hybrid Enrichment)
- Enrichment process is not 100% efficient.
- What is in the by-catch?

The Effect of Enrichment Efficiency and Genome Size on Expected Coverage



Parameters:

Mito Genome Size: 16KB

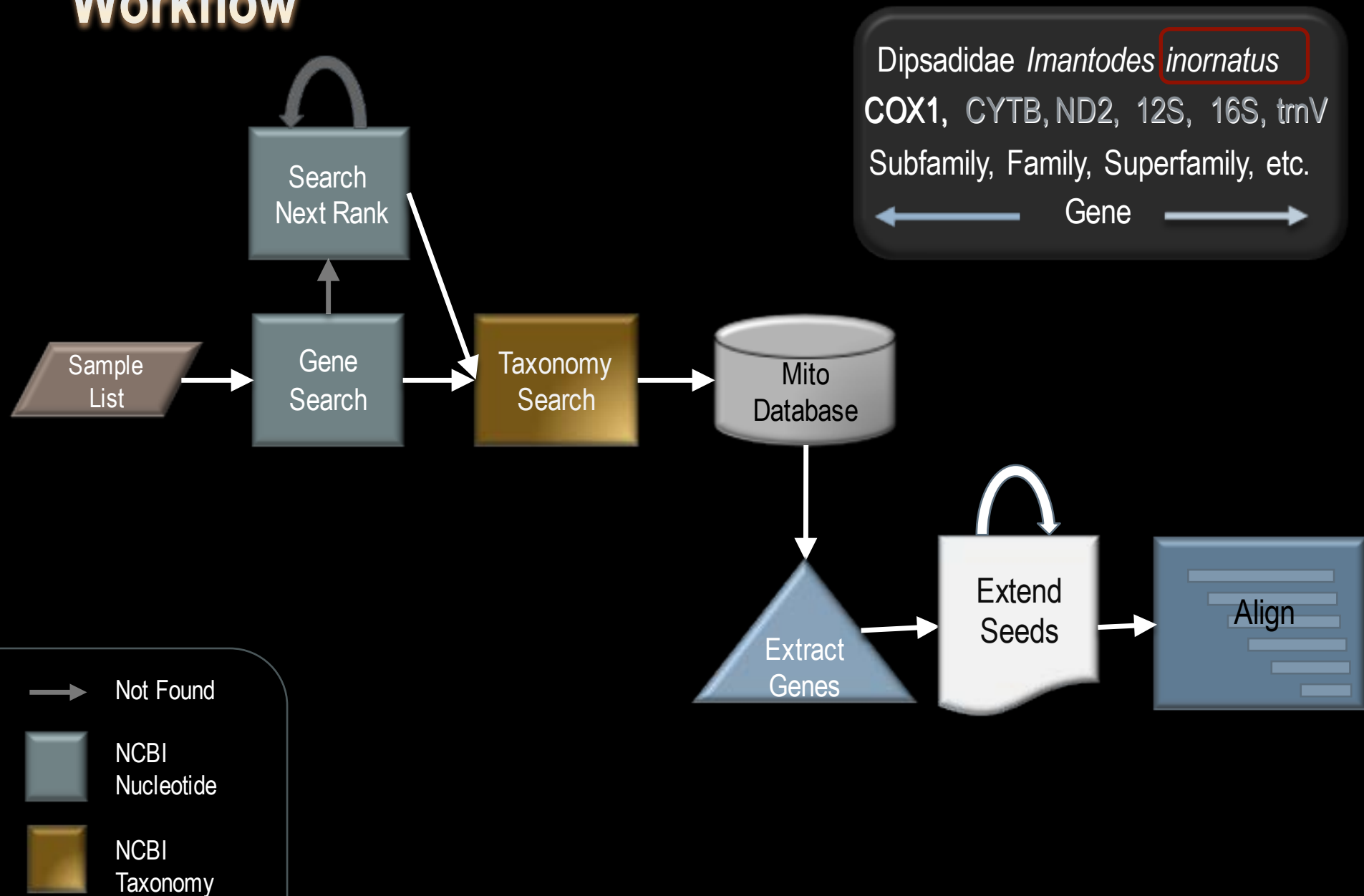
Mito Copy Number: 1000

Target Size: 100KB

The Problem

- De-Novo assembly too costly
- Reference genomes likely too distant for a complete reference based assembly (< 3% of samples have a full mitochondrial genome reference available at the genus or species level).
- Can use both (extending gene references) but there are several time consuming steps involved.
- Can automate the search for closely related genes and the process of extending genes.

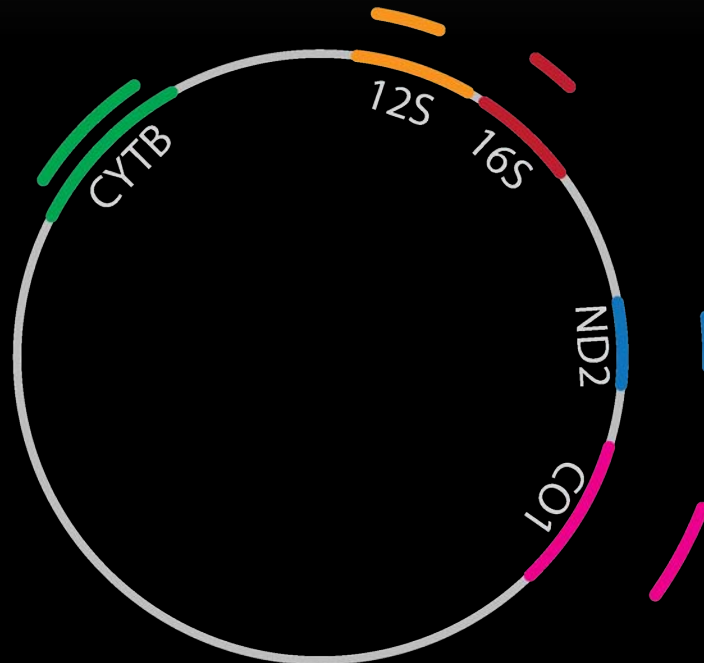
Workflow



Gene Extension



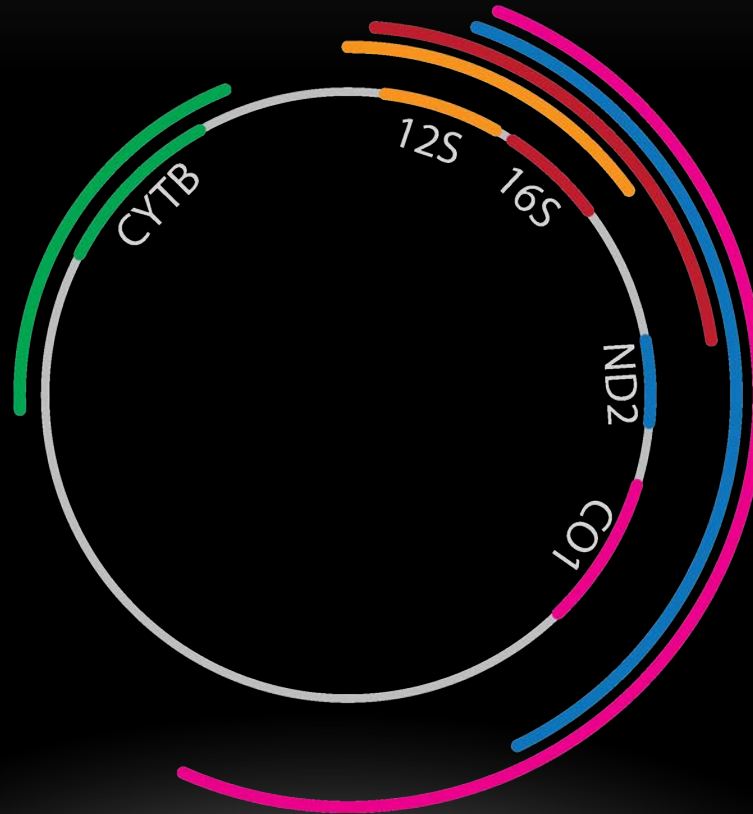
Gene Extension



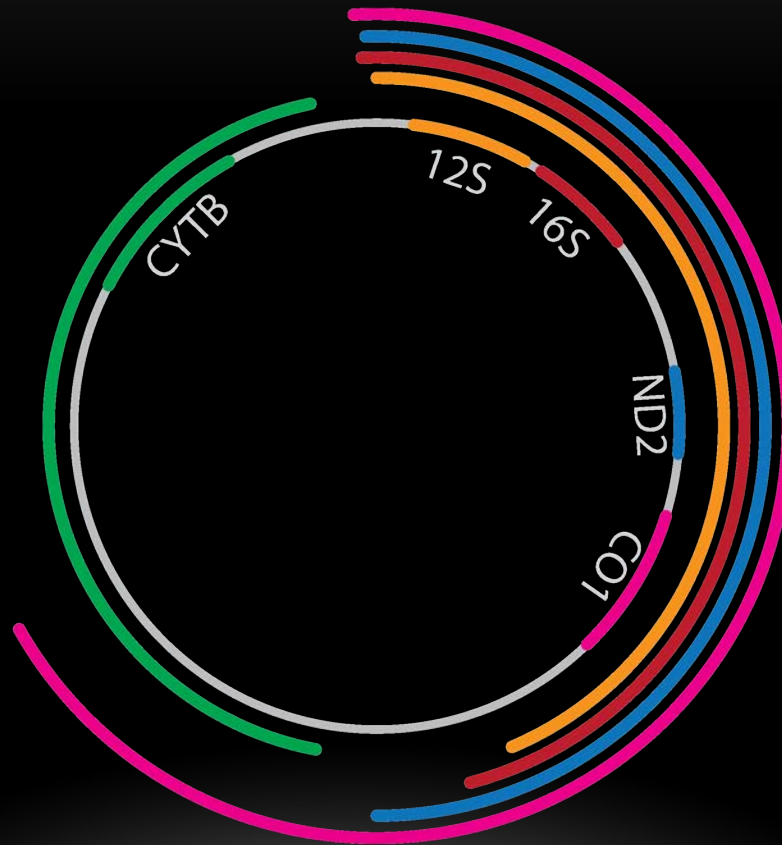
Gene Extension



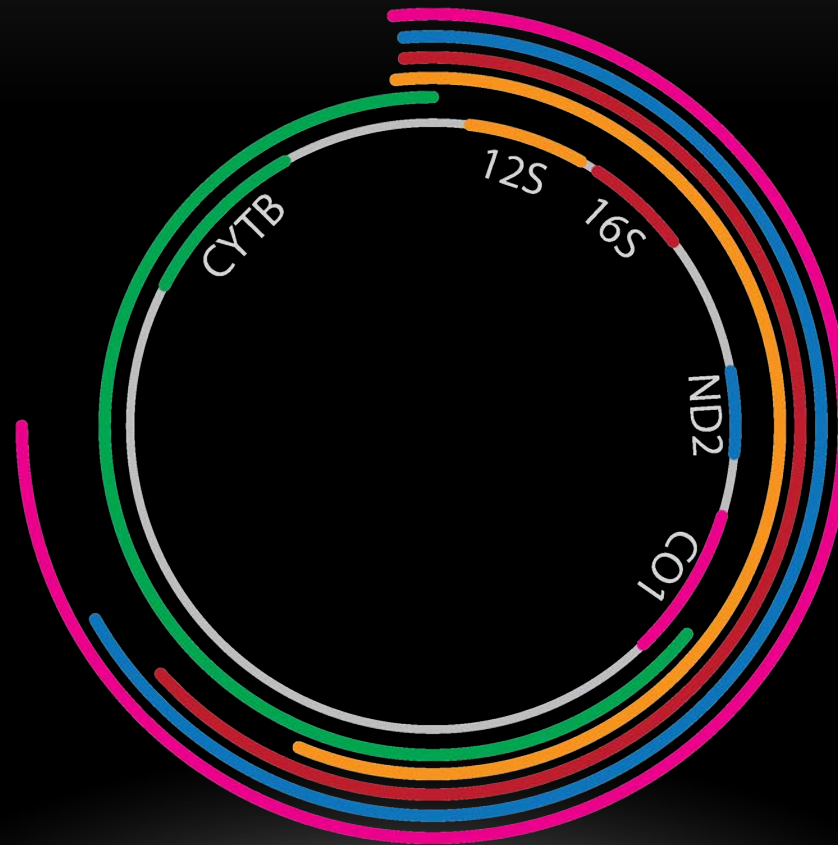
Gene Extension



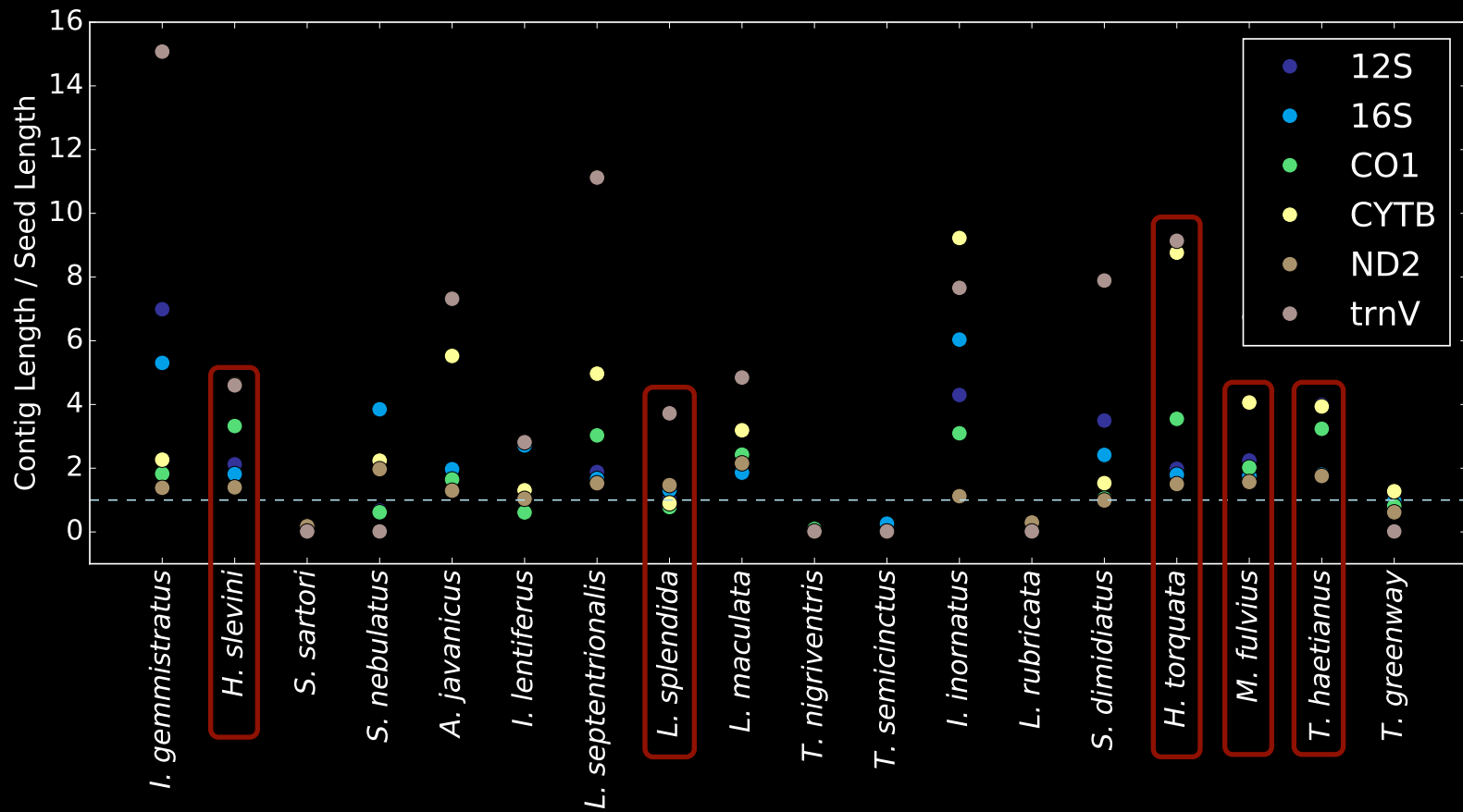
Gene Extension



Gene Extension



Ratio of Extended Seed (Contig) Length to Reference Length for 6 Mitochondrial Genes



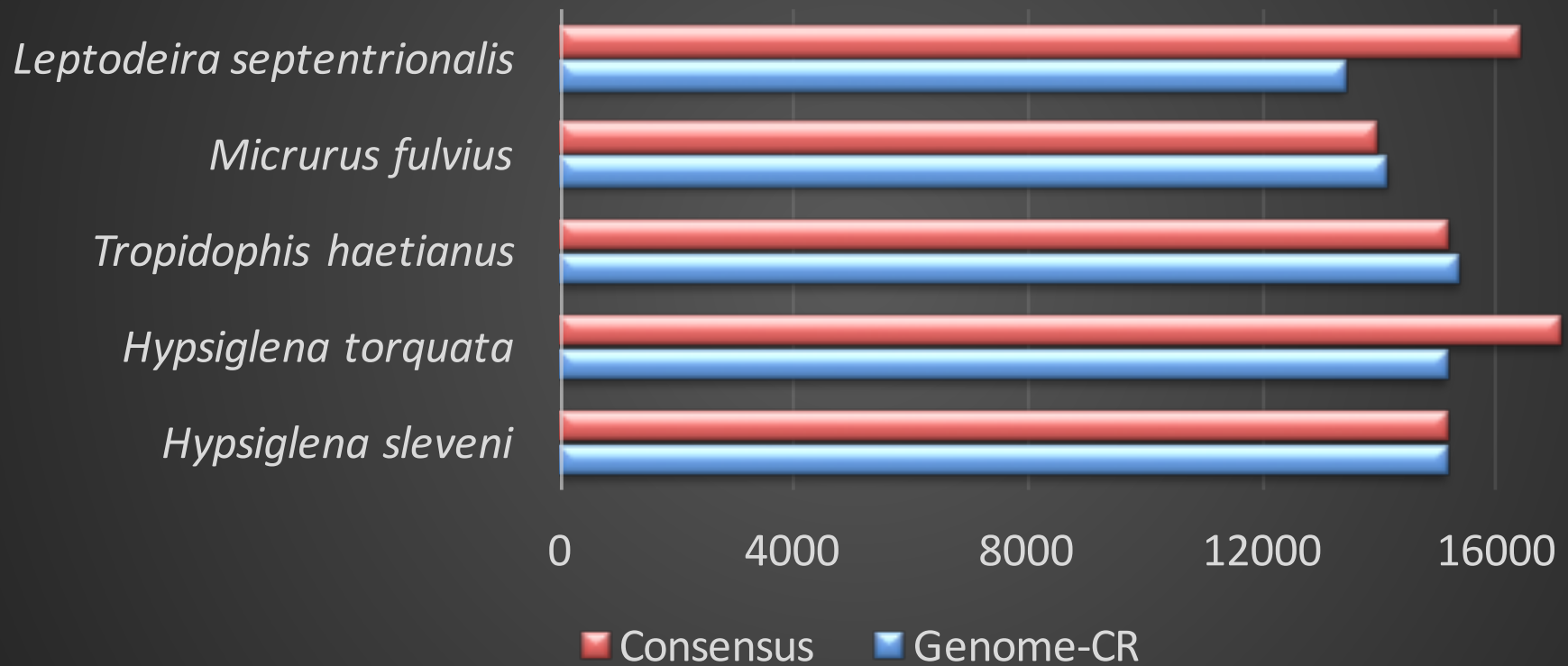
Results

Genome and Consensus Size in Nucleotides



Results

Reference without Control Region Compared to Consensus in Nucleotides



Future Directions

- Improve results by extending genes individually to avoid competition for reads
- Compare/modify the current extending algorithms to include iterative gap closure
- Apply to chloroplasts

Acknowledgements

- Alan Lemmon
- Emily Lemmon
- Lisa Barrow
- Sarah Banker
- Matt Mechtley

