Analysis of Phylogenetic Tree Methods on Chloroplast Genes

William Zhang
Project Outline

- Create a MSA on chloroplast dataset using MUSCLE
- Use RAxML, FastTree, PAUP* NJ to create trees from MSA
- Compare tree accuracy and runtime
Tree methods

- RAxML v8.2.0.10
  - Maximum likelihood based
- FastTree v2.1.9
  - Neighbor joining + maximum likelihood
- PAUP* v4a152
  - Neighbor joining using JC69 model
- What is fastest? Most accurate?
Gene selection

- **Gene 1: ndhF**
  - NADH dehydrogenase F
  - Often used to infer plant phylogeny

- **Gene 2: rbcL**
  - RuBisCO large subunit

- **Gene 3: atpA**
  - ATP synthase subunit alpha
  - More common for bacterial studies
Gene selection

- Intragenic spacer
  - Evolve very fast
  - psbA-trnH

- tRNA
  - Very highly conserved

- How does choice of dataset affect each method?
  - Highly conserved vs. not conserved
  - Correlation with whole genome studies
Bibliography


