

Analysis of Phylogenetic Tree Methods on Chloroplast Genes

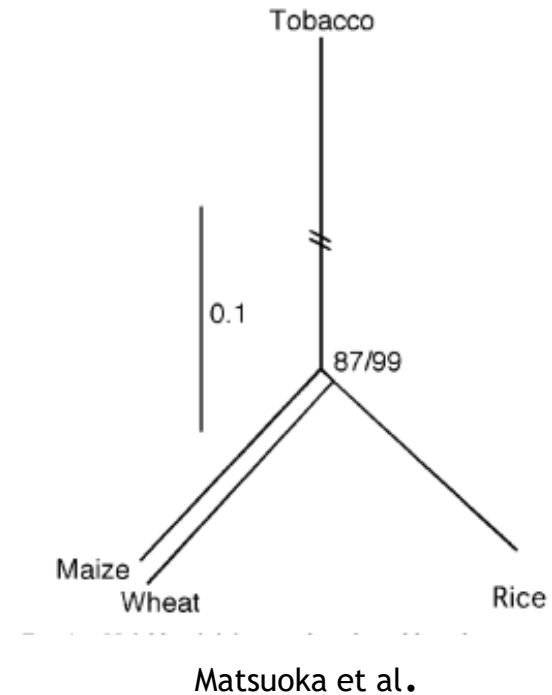
William Zhang

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

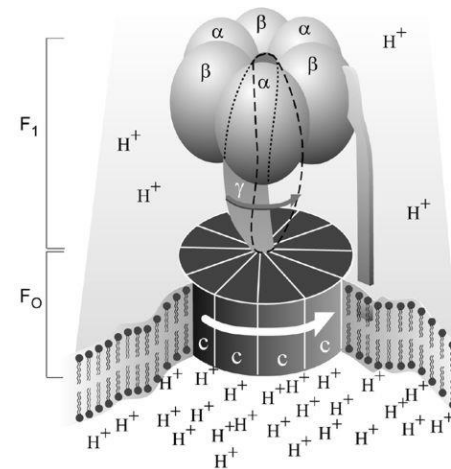
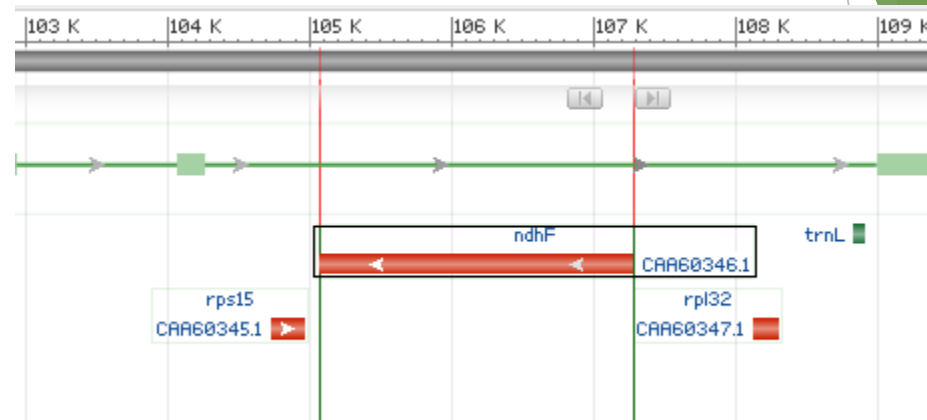
Tree methods

- ▶ 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

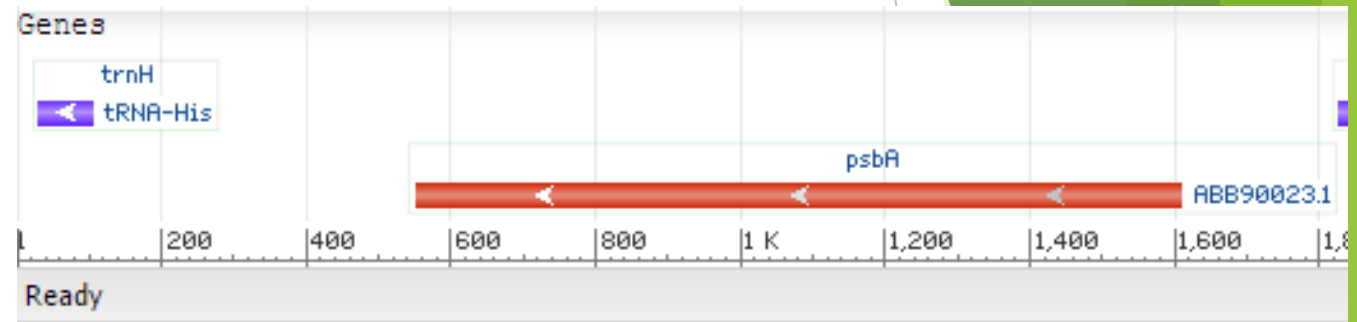
- ▶ Gene 1: *ndhF*
 - ▶ NADH dehydrogenase F
 - ▶ Often used to infer plant phylogeny
- ▶ Gene 2: *rbcL*
 - ▶ RuBisCO large subunit
- ▶ Gene 3: *atpA*
 - ▶ ATP synthase subunit alpha
 - ▶ More common for bacterial studies



Nobelprize.org

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

- ▶ K. X. Mayer, A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome *Science* 18 Jul 2014: Vol. 345, Issue 6194, 1251788 DOI: 10.1126/science.1251788 <http://science.sciencemag.org/content/345/6194/1251788.full>
- ▶ Gornicki, P., Zhu, H., Wang, J., Challa, G. S., Zhang, Z., Gill, B. S. and Li, W. (2014), The chloroplast view of the evolution of polyploid wheat. *New Phytologist*, 204: 704-714. doi:10.1111/nph.12931
- ▶ Yoshihiro Matsuoka, Yukiko Yamazaki, Yasunari Ogihara, Koichiro Tsunewaki; Whole Chloroplast Genome Comparison of Rice, Maize, and Wheat: Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. *Mol Biol Evol* 2000; 19 (12): 2084-2091. doi: 10.1093/oxfordjournals.molbev.a004033
- ▶ Degtjareva, G.V., Logacheva, M.D., Samigullin, T.H. et al. *Biochemistry Moscow* (2012) 77: 1056. doi:10.1134/S0006297912090131
- ▶ L Gielly, P Taberlet; The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences.. *Mol Biol Evol* 1994; 11 (5): 769-777. doi:10.1093/oxfordjournals.molbev.a040157
- ▶ Liu K, Linder CR, Warnow T (2011) RAXML and FastTree: Comparing Two Methods for Large-Scale Maximum Likelihood Phylogeny Estimation. *PLOS ONE* 6(11): e27731. <https://doi.org/10.1371/journal.pone.0027731>
- ▶ Hao C, Huang B, Yang L, Phylogenetic Relationships of the Genus *Taxus* Inferred from Chloroplast Intergenic Spacer and Nuclear Coding DNA, *Biological and Pharmaceutical Bulletin* Vol. 31 (2008) No. 2 P 260-265, <http://doi.org/10.1248/bpb.31.260>
- ▶ J Doyle, J Ballenger, E Dickson, T Kajita, and H Ohashi A, phylogeny of the chloroplast gene rbcL in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation., *Am. J. Bot.* April 1997 84:541

Bibliography (cont.)

- ▶ Naser S, Thompson FL, Hoste B, et al. Phylogeny and Identification of Enterococci by *atpA* Gene Sequence Analysis. *Journal of Clinical Microbiology*. 2005;43(5):2224-2230. doi:10.1128/JCM.43.5.2224-2230.2005.
- ▶ Patterson, T. B. and Givnish, T. J. (2002), PHYLOGENY, CONCERTED CONVERGENCE, AND PHYLOGENETIC NICHE CONSERVATISM IN THE CORE LILIALES: INSIGHTS FROM *rbcL* AND *ndhF* SEQUENCE DATA. *Evolution*, 56: 233-252.
- ▶ Kim, SC., Crawford, D.J., Jansen, R.K. et al. *Pl Syst Evol* (1999) 215: 85. doi:10.1007/BF00984649
- ▶ Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 -- Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, 5(3):e9490. doi:10.1371/journal.pone.0009490
- ▶ Alexandros Stamatakis; RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006; 22 (21): 2688-2690. doi: 10.1093/bioinformatics/btl446