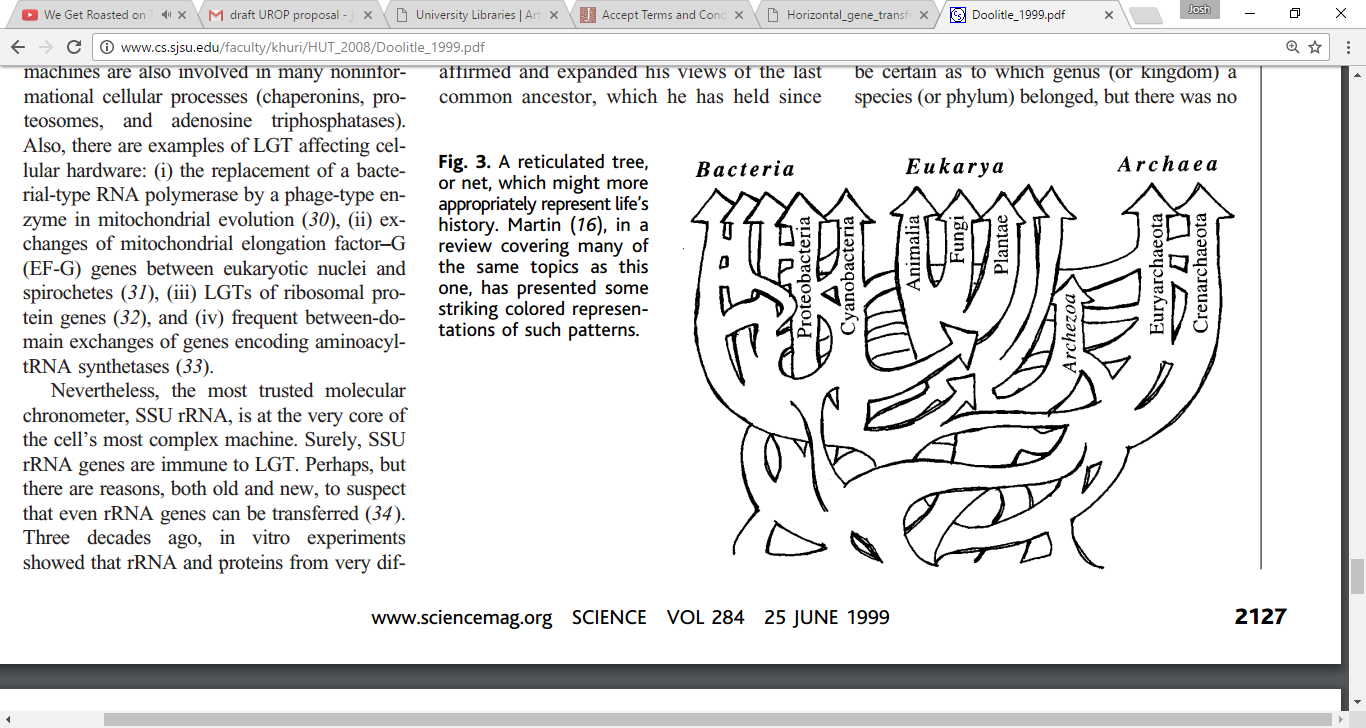
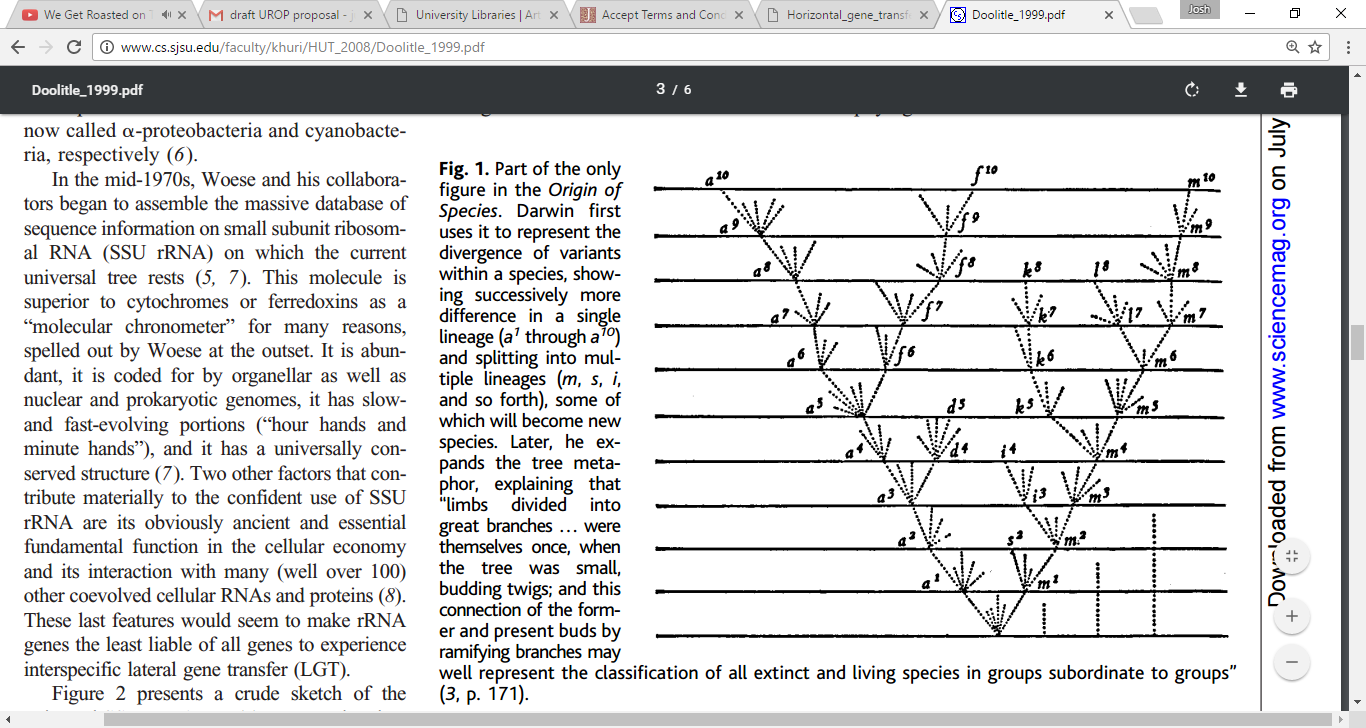
Disentangling the Effects of Hybridization on Macroevolutionary Analyses

Joshua Justison

Phylogenies are often constructed to gain some insight as to which species are most closely related or how far back species diverged from a most recent common ancestor. Numerous algorithms and methods have been developed to efficiently and precisely construct an estimated tree3. The end result is often a tree where species diverge from a common ancestor, branching off of each other (Figure 1A). These constructed phylogenies may then be used further when modelling trait evolution in order make inferences on how certain traits may have changed over time. However, in reality evolutionary relationships are more complex than a single ancestor node into two other lineages; lateral gene transfers in phylogenies and recombination and migration of species in gene trees would cause the mapping of evolutionary relationships to be more of a web of interactions between ancestors as opposed to just the standard tree1,4 (Figure 1B). Currently it is not known how much these hybridization events impact the modelling and estimation of trait evolution. An understanding to the adequacy of current phylogeny construction and trait evolution modelling is required to make accurate inferences about how traits evolve in species.



B

A

Figure 1. Classical interpretation of a phylogenetic tree and tree depicting interactions along the tree. A) Standard depiction of a phylogenetic tree taken from *The Origin of Species*8. A lineage here starts to diverge and eventually splits and branches out into different species. B) A phylogenetic tree depicting how some lineages may interact with other distant lineages, causing a tree that doesn’t directly make branches from other roots exclusively9.

The hypothesis is that hybridization events do have an impact on how well models predict trait evolution models, causing a lineage that has hybridization events to predict trait evolution values with a lower certainty and more bias than the predicted trait values of a tree without any hybridization. In this project I will work with Professor Goldberg to study the effects of how well current modelling can capture the complex relationships that hybridization events cause lineages to have and how well those models can make inferences about the lineages. The Goldberg lab conducts phylogenetic analyses by using phylogenetic models of trait evolution, utilizing many similar tools and methodology needed to explore the impacts of hybridization events.

To investigate the hypothesis, I will begin by choosing a history of the relationship between some species with known times of hybridization. That history will be simulated using msHOT to generate genetic variation data in the form of nucleotide sequences6. msHOT generates samples using a coalescent approach and has been modified from ms framework to allow for crossover hotspots, representing hybridization events5,6. Phenotype trait data will then also be simulated according to the history6. Since the samples are being simulated, we will be able to know the true rate of phenotype change. Next, a phylogenetic tree will be inferred by using generated genetic data. The trees will be constructed by using Bayesian methods as denoted by the software package, MrBayes10. Both the phenotype trait data and the phylogeny will be fit to a model as outlined by the Diversitree package in R to create a likelihood function7. Finally, a Bayesian analysis will be used on the likelihood function to create a probable surface to estimate the rate of the phenotypic change. The estimated surface can then be compared to the known values to investigate whether hybridization events impose any bias or uncertainty on parameter estimations.

The result of this project will shed some light into how well current modelling can capture the complex evolutionary relationships between species. If the effects of hybridization events impact the trait analysis, then we will be able to determine the strength of the effect and under which conditions the effect is stronger. The results will be presented at Undergraduate Symposium as a poster, or be written in a full-length scientific report. The project itself will have a budget of zero dollars as the software and hardware needed are publicly available. I hope this project will contribute to the disentanglement of signals in trait evolution analysis.

References

1Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, *7*(1), 44.

2Yuan, X., Miller, D. J., Zhang, J., Herrington, D., & Wang, Y. (2012). An overview of population genetic data simulation. *Journal of Computational Biology*, *19*(1), 42-54.

3Holder, M., & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature reviews genetics*, *4*(4), 275-284.

4Bapteste, E., Boucher, Y., Leigh, J., & Doolittle, W. F. (2004). Phylogenetic reconstruction and lateral gene transfer. *Trends in microbiology*, *12*(9), 406-411.

5Hellenthal, G., & Stephens, M. (2007). msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics*, *23*(4), 520-521.

6Hudson, R. R. (2004). ms a program for generating samples under neutral models.

7FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, *3*(6), 1084-1092.

8 C. Darwin, The Origin of Species by Means of Natural Selection (Murray, London, 1859).

9 Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science*, *284*(5423), 2124-2128.

10 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574.