Introduction to Phylogenetics & Tree Building Annotated Biblography

Phyloseminar 2024

Gavin Munson

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. <https://doi.org/10.1016/S0022-2836(05)80360-2>

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>

Higgins, D. G., & Sharp, P. M. (1988). CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene*, *73*(1), 237–244. <https://doi.org/10.1016/0378-1119(88)90330-7>

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>

I’m grouping the above papers together. I don’t think they all required their own description. They contain foundational documentation for the various alignment software’s. Each has its own advantages and shortcomings, and each discusses the algorithm and features available within each software. The reference sections are a rabbit hole of papers discussing the overarching theory behind the mechanisms of the models at work. Worth looking into if you need to know more about which specific program you decide to use for a project.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368–376. <https://doi.org/10.1007/BF01734359>

Felsenstein is a major character in the phylogenetic field. He is a major innovator and introduced many of the methods involved with phylogenetic influence. This paper is one of the first that introduces ML methods and their advantages over traditional parsimony.

Felsenstein, J. (1983). Parsimony in Systematics: Biological and Statistical Issues. *Annual Review of Ecology and Systematics*, *14*(1), 313–333. <https://doi.org/10.1146/annurev.es.14.110183.001525>

This paper discusses parsimony origins and major methods involved. It details the biological assumptions that parsimony applies and some of the philosophical issues related to using it.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*(4), 783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>

Felsenstein introduces the use of bootstrapping for phylogenetic inference here. It provides a practical method for placing confidence intervals on phylogenetic estimates, such as the use of consensus trees constructed from larger simulated bootstrap samples.

Fitch, W. M., & Margoliash, E. (1967). Construction of Phylogenetic Trees. *Science*, *155*(3760), 279–284. <https://doi.org/10.1126/science.155.3760.279>

An early introduction into the how to build phylogenetic trees based on mutations for accurate classification. The authors are noticing that data can produce imperfect trees due to the statistical analysis criterion.

Fitch, W. M., & Markowitz, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochemical Genetics*, *4*(5), 579–593. <https://doi.org/10.1007/BF00486096>

This paper analyses codon variability and starts to apply methods for studying fixation and mutation rates within nucleotides. Some key take-aways are that rates are variable for different genes, and that most mutations are selectively neutral.

Koonin, E. V. (2005). Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics*, *39*(1), 309–338. <https://doi.org/10.1146/annurev.genet.39.073003.114725>

A decent review paper that covers the different types of homologous genes; Orthologs reflecting evolutionary relationships between species, paralogs arising from a gene duplication event, and xenologs which are acquired through horizontal gene transfer.

Le, S. Q., Dang, C. C., & Gascuel, O. (2012). Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Molecular Biology and Evolution*, *29*(10), 2921–2936. <https://doi.org/10.1093/molbev/mss112>

One of many example papers that investigated the use of different substitution matrices for different site evolutionary rates. One of the mixture model systems being used in practice.

Lockhart, P. J., Steel, M. A., Hendy, M. D., & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution*, *11*(4), 605–612. <https://doi.org/10.1093/oxfordjournals.molbev.a040136>

Investigated assumptions of nucleotide compositional homogeneity and the improper species groups associated with it. Argues that irregular nucleotide composition can create misleading tree reconstructions.

Philippe, H., & Telford, M. J. (2006). Large-scale sequencing and the new animal phylogeny. *Trends in Ecology & Evolution*, *21*(11), 614–620. <https://doi.org/10.1016/j.tree.2006.08.004>

This paper address stochastic errors with small datasets and proposes a need for large scale sequencing of many taxa and many genes to reduce errors in tree reconstruction.

Rannala, B., & Yang, Z. (1996). Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution*, *43*(3), 304–311. <https://doi.org/10.1007/BF02338839>

An early example of introducing Bayesian inference into character-based phylogenetic inference. Uses a MCMC model process and introduces maximum posterior probability (MAP) tree reconstruction. Find that the topologies are similar to that of ML, but with very different posterior probabilities compared to bootstrap values.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>

This paper introduces the neighbor-joining method. This is now one of the most popular distance-based methods for tree making.