Following decades-long work of many, we will study sequence motif analysis.

1. **Modeling motifs**. Read Stormo [1]'s brief history of computational motif analysis.
   (a) On canvas, you will find a fasta file containing thousands of known 9bp motifs of a transcription factor. Estimate the parameters of an independent sites model for the motif.
   (b) Read the Markov chain notes through the section on Likelihood. Does a *nonhomogeneous* Markov chain fit the motif better than the independence model of Part a? Compute a p-value using:
      i. theory and,
      ii. a Monte Carlo method.
   (c) The Markov chain allows for dependence between random variables, but it is of a very specific form and there are many other kinds of dependence. Can you find evidence of other kinds of dependence beyond what you might have already discovered in Part b? Provide numeric measures of confidence for your conclusions.
2. **Testing for known motif enrichment.**
3. **Detecting novel motifs.**

References

[1] Gary D. Stormo. "DNA binding sites: representation and discovery". In: *Bioinformatics* 16.1 (2000), pp. 16–23. ISSN: 1367-4803. URL: https://dx.doi.org/10.1093/bioinformatics/16.1.16.