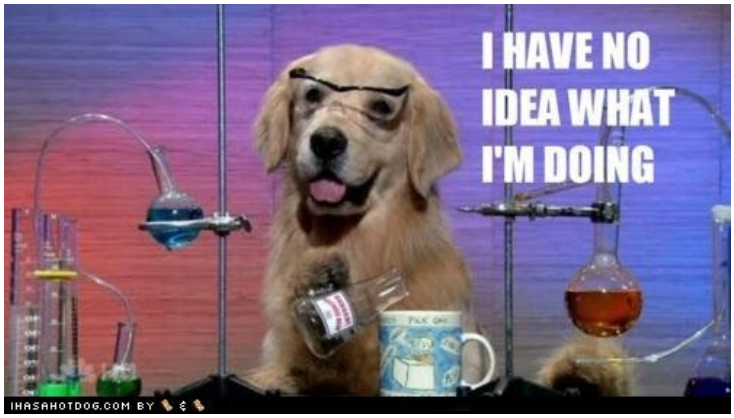
Evil Nature Paper :-(

Shaun Barker, Alex Jackson, Jessica Parker

November 25, 2015



- Start with 2 lots of DNA **(alleles?)**
- Divide sequence into 100 bp bins.
- Assign each bin either missing (.), heterozygous (1) or homozygous (0). The sequence of this gives the input to the HMM e.g. $(0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, \dots)$.
- Divide time into $0 = t_1 < t_2 < \cdots < t_n = T_{\max} < t_{n+1} = \infty$. Choose $T_{\max}$ such that only a few percent of coalescent events fall in $[T_{\max}, \infty)$. **How do you choose this?**
- Hidden states: which interval $k = [t_k, t_{k+1})$, the TMRCA/coalescent times for the different bins fall into. **Figure 1** on the paper is quite useful for illustrating this.
- METHOD...
- Assume constant population size.
- Observations into HMM, find TMRCAs of the bins and associated parameters e.g. hidden states,
- Evaluate the likelihood of the sequence using EM (expectation maximisation).
- Powell's direction set: Got a discretised $\lambda(t)$ function (piecewise constant), optimise one "step" at a time. **???**
- Iterate 20 times.