

Arpit Tandon

Lecture13

All general definitions and formulae from lecture 12. Using that basic information to derive more:

Probability of having distinct ancestor for n sequences

- $P(n) = 1 - \frac{\binom{n}{2}}{2N}$

Probability of having coalescent at generation t+1

- $P(n)t = \frac{\binom{n}{2}}{2N} e^{-\frac{\binom{n}{2}}{2N}t}$

Coalescence time

- $E(T_n) = \frac{2N}{\binom{n}{2}}$

Total Length

- $E(T_{tot}) = 4N \sum 1/i$

Expected number of mutations

- $E(S) = 4N\mu \sum 1/i$

Where μ is the mutation we get from one generation to next

Inference about μ

This can be estimated from S

- $4N\mu = S/\sum 1/i$, with this equation, we can estimate either N or μ , given other

With the help of these equations and simulated coalescent model, we can place mutation in genealogy. Most often, they fall in longer branches. Even by looking at them, one can predict how many pair of sequences will be affected.

Repetition of simulation model gives a predicted mutation spectrum. See slide 18. These spectrum shows the mutation occurs more often at bottom part in dense branches.

Number of mutations can be derived from coalescent model by tracing back in time until we reach MRCA, keeping track of mutation events.

Technique to do this

Assuming no recombination

- Take probability of a coalescent event (depends on population size) P_{ca}
- Take probability of a mutation (depends on mutation rate) P_{mut}
- Proceed backward in coalescent model
- For 2 identical sequences
- $P_2(S \text{ is } 0) = \frac{P_{ca}}{P_{ca} + P_{mut}}$
 - $= \frac{1}{1 + \Theta}$
- Probability that first j events are mutation
- $= P_2(j) = \left(\frac{\Theta}{1 + \Theta}\right)^j \frac{1}{1 + \Theta}$
- For n sequences
- $P_n(j) = \sum p_{n-1}(j-1) Q_n(i)$