

Arpit Tandon  
Lecture 12

### **Coalescent Models**

A Retrospective model of population genetics to trace all alleles of a gene shared by all members of given population to a single ancestral copy, MRCA.  
This allows us to figure out the relatedness between sequences.

### **Factors of genetic variation**

- Number of sequences
- mutation rate
- Sequence length

### **Refer Lecture 12, slide 8**

Time gap can be figured out building up a tree joining two closest one to reach their MRCA.

A gene tree shows which sampled sequences are most closely related to each other and perhaps the times when the MRCA

These tree structure helps us in identifying the allele combination.

If subtrees are disjoint, there cant be a pair of mutated allele. If mutations occur within a subtree, then we can have pair of Mutated allele.

### **Mutation model**

Infinite allele----every mutation forms different allele

Infinite site ----every mutation occurs at a different site

### **Assumptions to infer**

- no recurrent mutation occur

### **Parameters to focus on**

- Mutation rate
- Population size
- Time
- Sample size
- Recombination rate

Probability of two sequences having same ancestor =  $1/2N$

Probability of two sequences having distinct ancestors =  $1-1/2N$

The above statement if for population size of  $2N$  and for only 1 previous generation.

For  $t+1$  generation or time.

Probability of two sequences having a MRCA at time  $t+1 = 1/2N(1-1/2N)^t$

Probability of  $n$  sequences having  $n$  distinct ancestors =  $\prod (1 - i/2N)$

For  $t+1$  generation or time

Probability of  $n$  sequences having a MRCA at time  $t+1 = \prod (i/2N)(1-i/2N)^t$

**Refer slides 21 and 24**

Using exponential distribution, we deduce the following

Decay rate =  $\prod i/2N$

Mean =  $\prod 2N/i$

Total time =  $4N \sum 1/i$

Expected number of mutations =  $4N\mu \sum 1/i$

$\theta = 4N\mu$  describes the mutation rate between generations

Expected number of mutation increases as the size increases. Also, variance increases too and there will be maximum variance when  $n$  is small which shows the maximum variance occurs due to early coalescence.

Assuming no recurrent mutation

How to infer about  $\theta$

**Refer slides 31- 35 for reference.**

$\text{Var}(\theta)$  is inversely proportional to  $n$  means as  $n$  increases mutation decreases, thus decreasing the variance.