# **polyARG Maths**

Although these are closely linked concepts, it is important to distinguish that we are no longer trying to derive identity by descent (or the probability that the two individuals inherited an allele from the same parent). We are now looking at when two alleles have coalesced (e.g. TMRCA).

## **Single Loci, Standard Coalescent**

For our initial condition, we will consider loci unlinked and free from recombination (recombination rate = 0). In addition, we are assuming the underlying process was generated by a basic Wright-Fisher model and diploid organism.

**Givens**

Under basic coalescent theory, we can assume:

1. Probability of selecting two of the same allele at each time-point is:
2. Probability of not selecting the same allele at each time-point is:
3. Therefore, expected time to coalescence is just the inverse of prob of coalesce in single generation (mean of geometric is 1/p):

As a result, the probability that two alleles coalesce in the previous generation (geometric distribution):

*These are non-parametric, straightforward realizations of the coalescent model.*

**Coalescent Likelihood**

**Extension to Multiplicity of Infection**

Now, let’s assume that the number of chromosomes is a random variable (i.e. the number of strains in a sample is a RV).

n(n-1)

## **Two-Locus Ancestral Recombination Graph**

We are now going to simulate a Wright-Fisher population of N individuals with two loci. However, each time a parent is sampled, the probability that the chromosome remains intact is 1- ρ. When recombination occurs, the homologous loci of that parent will be inherited.

**Givens**

1. Prob

These models all start with Feinstein 1981