

II. Population Genetics

Coalescent Models with Recombination

Lecture 14

Instructor: Su-In Lee

Last Lectures

- Basic Properties of the Coalescent
 - MRCA
 - Coalescence times
 - Number of mutations
- Frequency spectrum of polymorphisms
- Predicting number of variants in a sample

Today ...

- Further refining the coalescent
 - Recombination
 - Migration
- Discussion of potential applications

Recombination ...

- No recombination
 - Single genealogy
- Free recombination
 - Two independent genealogies
 - Same population history
- Intermediate case
 - Correlated genealogies

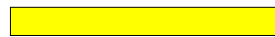
The History of Two Sequences

- Let's consider the potential history of two sequences, but this time... with a twist!

Sequence A

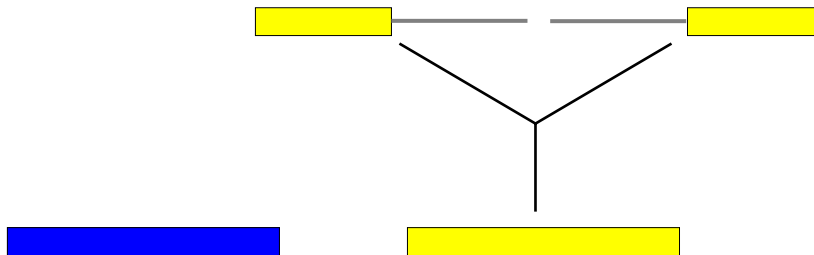


Sequence B



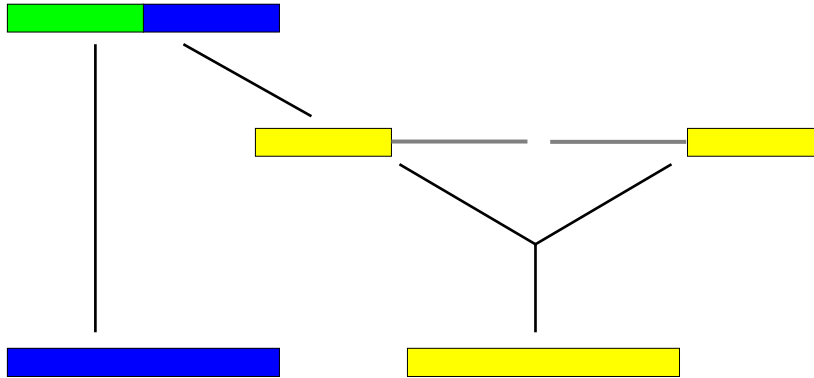
The History of Two Sequences

- Before we reach a common ancestor ... we find that sequence B is actually the result of recombination between two ancestral sequences



The History of Two Sequences

- The next event we encounter is a coalescence event, as expected ...

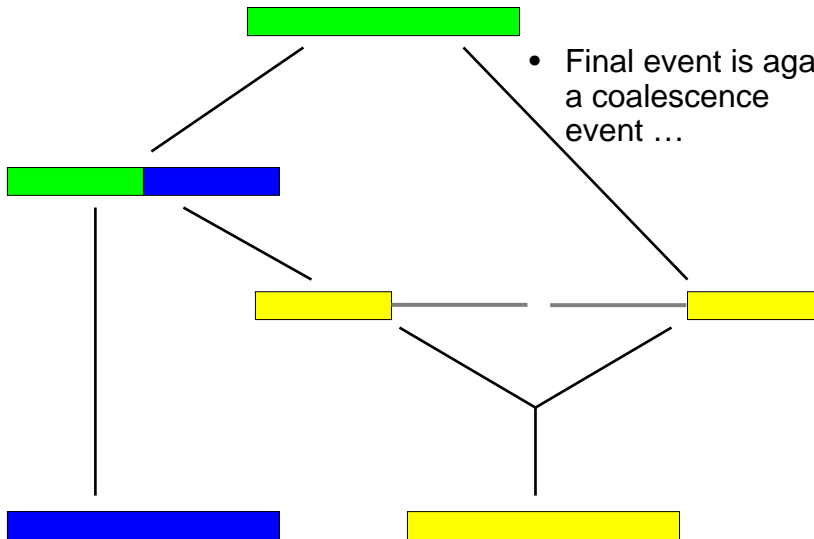


Computational Molecular Biology and Genomics, Spring 2009

7

The History of Two Sequences

- Final event is again a coalescence event ...



Computational Molecular Biology and Genomics, Spring 2009

8

Potential Consequences ...

- Different portions of the sequence have different coalescence times
- Different portions of the sequence will show more or less variation

Simulating the Coalescent with Recomb

- Assume the various alternative events are rare
- Time until the next event is approximately exponentially distributed
- Conditional on something happening, figure out whether it was:
 - Recombination
 - Coalescence

Generating Genealogies

- Proceed backwards in time, until...

- Coalescent event

- Reduces number of ancestors by 1

$$P_{CA} \approx \binom{n}{2} / 2N$$

- Recombination

- May increase number of ancestors by 1

$$P_{rec} \approx nr$$

P(First Event is CA)

$$\begin{aligned} P(\text{no rec}) &= \frac{P_{CA}}{P_{CA} + P_{rec}} = \frac{\binom{n}{2} / 2N}{\binom{n}{2} / 2N + nr} \\ &= \frac{n-1}{4Nr + n-1} \\ &= \frac{n-1}{R + n-1} \end{aligned}$$

Coalescent with Recombination

- Analytical results are difficult
- Typical approach is to ...
 - First, simulate ancestral recombination graphs (ARG)
 - Coalescent tree with recombination events
 - Study sample properties implied by simulated ARGs
 - For example, similarity in frequencies of neighboring SNPs

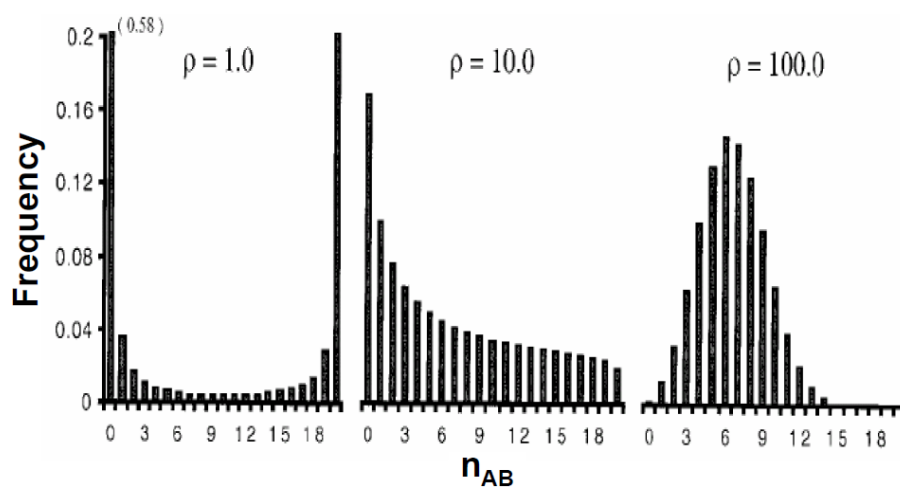
Correlated Genealogies

- Produce correlation in
 - Allele frequencies
 - Number of mutations
 - Distribution of alleles among chromosomes
 - Linkage disequilibrium
- Use simulations to evaluate distributions as a function of recombination rate

Example 1

- Consider a sample of $n = 90$ chromosomes
- 2 locus coalescent, focus on samples where
 - $n_A = 30$
 - $n_B = 20$
- What is the distribution of n_{AB} ?
 - And consequently of D' , r^2

Impact of Recombination on Haplotype Distribution



Some Notes ...

- If we are interested in studying the local recombination rate, neither r^2 or D' retain all the information contained in n_A , n_B , n_{AB}
- We can estimate R or ρ by finding the value that maximizes the probability of the observed sample configuration

Estimating Recombination Rates

- McVean et al. (*Science*, 2004) estimated the following “pseudo-likelihood” for a sample of haplotypes:

$$l(4Nr) = \sum_{i,j} l(n_i, n_j, n_{ij} | 4Nr_{ij})$$

Summations is over all pairs of markers

- Estimated recombination rates allow us to predict what other chromosomes or samples from the population might look like.

Recommended Reading

- **Richard R. Hudson (1990)**
- *Gene genealogies and the coalescent process*
- Oxford Surveys in Evolutionary Biology, Vol. 7. D. Futuyma and J. Antonovics (Eds). Oxford University Press, New York.

Acknowledgement

- This lecture note is based on material by Profs Goncalo Abecasis (Univ of Michigan) and Itsik Pe'er (Columbia Univ).
- Chapter 3 of *Principles of Population Genetics* by Daniel L. Hartl & Andrew Clark