CSCI 1820 Notes

Scribes: fshi3, hni2

March 20

Chapter 3: Recombination and Ancestral Recombination Graphs Algorithms

Reconstructing genealogies for a set of individuals case-control study

\[ \begin{align*}
&\text{Cases Austin} \\
&\text{Controls Parents}
\end{align*} \]

We use GWAS-genome-wide association study to find the statistical associations between variations and traits.

SNP = single nucleotide polymorphisms

The 10,001th base:

\[
\begin{array}{c}
\text{AATAGGA} \\
\text{AATAGGA} \\
\text{AATGGGA} \\
\text{AATAGGA} \\
\text{AATGGGA} \\
\text{AATGGGA}
\end{array}
\]

Alleles at this SNP: A and G.

Infinite Allel Model: With large enough of population, every mutation will be unique.

**Coalescent Model & Recombination**

Phylogenetic Tree: Set of sequences.

\[
\begin{align*}
00100 & \quad 00100 \\
\downarrow & \quad \downarrow \\
00100 & 00110 & \Rightarrow & 00100 & 00100 \quad Coalescence
\end{align*}
\]
PhylogeneticTree = \{ \text{mutation coalesces} \}

ARG = \{ \text{mutation coalesces recombine} \}

Assuming C is a sequence of [0, 1, •]
C[i] is the ith SNP of C

DEF ’∼’:
\[ C_1[i] ∼ C_2[i] \text{ iff } \]
\[ C_1[i] = C_2[i] \]
\[ \text{or } C_1[i] = • \]
\[ \text{or } C_2[i] = • \]

DEF ’Γ’:
if \[ C[i] = 0 \] then \[ ΓC[i] = 1 \]
and vice versa.
\[ Γ• = • \]

TRACTS - Colored fragments. Recombination tracts

Shared Tract
There is a sheared tract between \( C_1 \) and \( C_2 \) with a contiguous set of markers \( a, ..., b \) if the following conditions hold:
1. \( C_1[i] ∼ C_2[i] \) for all \( a \leq i \leq b \)
2. There is at least one \( i \) for which: \( C_1[i] = C_2[i] \neq • \)
3. If \( a > 1 \) then \( C_1[a−1] \neq C_2[a−1] \) and neither of them is •
4. If \( b < m \) then \( C_1[b+1] \neq C_2[b+1] \) and neither of them is •