Last Time

- Population assignment examples
- Forensic evidence and individual identity
- Introduction to paternity analysis
Today

- Using $F_{ST}$ to estimate migration
- Direct estimates of migration: parentage analysis
- Introduction to phylogenetic analysis
Island Model of Population Structure

expected identity by descent at time $t$, no migration:

$$f_t = \frac{1}{2N} + (1 - \frac{1}{2N})f_{t-1}$$

IBD due to random mating

IBD due to prior inbreeding

Incorporating migration:

$$f_t = \left[\frac{1}{2N} + (1 - \frac{1}{2N})f_{t-1}\right](1 - m)^2$$

where $m$ is proportion of $N$ that are migrants each generation

If population size on islands is small, and/or gene flow ($m$) is low, drift will occur
Migration-Drift Equilibrium

- This approach is VERY widely used to calculate number of migrants per generation

- It is an APPROXIMATION of EQUILIBRIUM conditions under the ISLAND MODEL

- Only holds for low $Nm$
  - $F_{ST} = 0.01$, approximation is $Nm = 24.8$
  - If $N=50$, actual $Nm$ is 14.6

At migration-drift equilibrium:

$$f_t = \left[ \frac{1}{2N} + (1 - \frac{1}{2N}) f_{t-1} \right] (1-m)^2$$

Assuming $m^2$ is small, and ignoring $2m$ in numerator and denominator:

$$F_{ST} = \frac{(1-m)^2}{2N - (2N-1)(1-m)^2}$$

$$F_{ST} \approx \frac{1}{4Nm + 1} \quad Nm \approx \frac{1 - F_{ST}}{4F_{ST}}$$
Differentiation of Subpopulations

- Subpopulations will be more uniform with high levels of gene flow and/or high N
- $Nm>1$ homogenizes populations
- $Nm<<1$ results in fixation of alternate alleles and ultimate differentiation ($F_{ST}=1$)
Limitations of $F_{ST}$

- $F_{ST}$ is a long, integrated look into the evolutionary/ecological history of a population: may not represent status quo

- Assumptions of the model frequently violated:
  - Island model unrealistic
  - Selection is often an important factor
  - Mutation may not be negligible
  - Sampling error!
Alternatives to $F_{ST}$

- Direct measurements of movement: mark-recapture

- Genetic structure of paternal and maternal gametes only
  - Chloroplast and mitochondrial DNA
  - Pollen gametes

- Parentage analysis: direct determination of the parents of particular offspring through DNA fingerprinting
Paternity Exclusion Analysis

- Determine multilocus genotypes of all mothers, offspring, and potential fathers
- Determine paternal gamete by “subtracting” maternal genotype from that of each offspring.
- Infer paternity by comparing the multilocus genotype of all gametes to those of all potential males in the population
- Assign paternity if all potential males, except one, can be excluded on the basis of genetic incompatibility with the observed pollen gamete genotype
- Unsampled males must be considered
**Paternity Exclusion**

- First step is to determine paternal contribution based on seedling alleles that do not match mother

  - Notice for locus 3 both alleles match mother, so there are two potential paternal contributions

- Male 3 is the putative father because he is the only one that matches paternal contributions at all loci
Sweet Simulation of Paternity Analysis

- Collected seeds (baby) from a hermaphroditic, self-incompatible plant

- Which of the candidate hermaphrodites (you) fathered the seeds?

- Six loci with varying numbers of alleles
  1. Organize candy into loci (next slide)
  2. Determine paternal contribution to offspring by subtracting maternal alleles
  3. Exclude potential fathers that don’t have paternal allele
  4. Nonexcluded candidate is the father!
Loci
Mother and Child

Maternal Alleles

Paternal Alleles
Alleles versus Loci

- For a given number of alleles: one locus with many alleles provides more exclusion power than many loci with few alleles
  - 10 loci, 2 alleles, Pr = 0.875
  - 1 locus, 20 alleles, Pr=0.898

- Uniform allele frequencies provide more power
Characteristics of an ideal genetic marker for paternity analysis

- Highly polymorphic, (i.e. with many alleles)
- Codominant
- Reliable
- Low cost
- Easy to use for genotyping large numbers of individuals
- Mendelian or paternal inheritance
Shortcomings of Paternity Exclusion

- Requiring exact matches for potential fathers is excessively stringent
  - Mutation
  - Genotyping error

- Multiple males may match, but probability of match may differ substantially

- No built-in way to deal with cryptic gene flow: case when male matches, but unsampled male may also match
  - Type I error: wrong father assigned paternity)
Is it possible we're implicating the wrong father in our paternity exclusion analysis?

Look at mismatching loci and the genotypes. Could you have been wrongly excluded?
Probabilistic Approaches

- Consider the probability of alternative hypotheses given the data

- Probabilities are conditioned based on external evidence (prior probabilities)

\[
P(H_1 \mid E) = \frac{P(E \mid H_1)P(H_1)}{P(E \mid H_1)P(H_1) + P(E \mid H_2)P(H_2)}
\]
Likelihood Approach for Paternity Assignment

Consider two hypotheses:

- Alleged father is the true father
- A random male from the population is the true father

Calculate a score for each male, reflecting probability he is correct father:

\[
L(H_1, H_2 | g_m, g_a, g_o) = \frac{T(g_o | g_m, g_a).P(g_m).P(g_a)}{T(g_o | g_m).P(g_m).P(g_a)} = \frac{T(g_o | g_m, g_a)}{T(g_o | g_m)}
\]

where

- \(H_1\) is probability male \(a\) is father,
- \(H_2\) is probability male \(a\) is not father
- \(T\) is transition probability
- \(P\) is probability of observing the genotype

and \(g_o, g_m\) and \(g_a\) are genotypes of offspring, mother, and alleged father.
### Transition Probabilities

Table 1 Likelihood ratios for all compatible mother–alleged father–offspring trios. X represents any allele other than B; Y represents any allele that is neither B nor C. The frequencies of alleles B and C are denoted b and c. The likelihood ratio, \( L(H_1,H_2) \), is the probability of the offspring’s genotype given the mother’s and alleged father’s genotypes, \( T(g_o \mid g_m, g_a) \), divided by the probability of the offspring’s genotype given the mother’s genotype, \( T(g_o \mid g_m) \). A similar table is shown in more condensed form in Brenner (1997).

<table>
<thead>
<tr>
<th>Offspring’s genotype ( (g_o) )</th>
<th>Alleged father’s genotype ( (g_a) )</th>
<th>Mother’s genotype ( (g_m) )</th>
<th>( T(g_o \mid g_m, g_a) )</th>
<th>( T(g_o \mid g_m) )</th>
<th>( L(H_1,H_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>BB</td>
<td>BB</td>
<td>1</td>
<td>( b )</td>
<td>( 1/b )</td>
</tr>
<tr>
<td>BB</td>
<td>BX</td>
<td>BB</td>
<td>1/2</td>
<td>( b )</td>
<td>( 1/2b )</td>
</tr>
<tr>
<td>BB</td>
<td>BX</td>
<td>BX</td>
<td>1/2</td>
<td>( b/2 )</td>
<td>( 1/b )</td>
</tr>
<tr>
<td>BC</td>
<td>BB</td>
<td>CC</td>
<td>1</td>
<td>( b )</td>
<td>( 1/b )</td>
</tr>
<tr>
<td>BC</td>
<td>BX</td>
<td>CC</td>
<td>1/2</td>
<td>( b )</td>
<td>( 1/b )</td>
</tr>
<tr>
<td>BC</td>
<td>BB</td>
<td>CY</td>
<td>1/2</td>
<td>( b/2 )</td>
<td>( 1/b )</td>
</tr>
<tr>
<td>BC</td>
<td>BX</td>
<td>CY</td>
<td>1/4</td>
<td>( b/2 )</td>
<td>( 1/b )</td>
</tr>
<tr>
<td>BC</td>
<td>BB</td>
<td>BC</td>
<td>1/2</td>
<td>( (b + c)/2 )</td>
<td>( 1/(b + c) )</td>
</tr>
<tr>
<td>BC</td>
<td>BY</td>
<td>BC</td>
<td>1/4</td>
<td>( (b + c)/2 )</td>
<td>( 1/2(b + c) )</td>
</tr>
<tr>
<td>BC</td>
<td>BC</td>
<td>BC</td>
<td>1/2</td>
<td>( (b + c)/2 )</td>
<td>( 1/(b + c) )</td>
</tr>
</tbody>
</table>


\[
L(H_1, H_2 \mid g_m, g_a, g_o) = \frac{T(g_o \mid g_m, g_a)}{T(g_o \mid g_m)}
\]
LOD Score for Paternity: “Cervus” program

- Combine likelihoods across loci by multiplying together

\[ L = \prod_{i=1}^{m} L_i \]

- Calculate Log Odds Ratio (LOD)

\[ LOD = \ln(L) \]

- What is a significant LOD?
  - No good criteria
  - Use difference between most likely (LOD₁), next most likely (LOD₂)

\[ \Delta = LOD_1 - LOD_2 \]

Other programs listed at NIST website: http://www.cstl.nist.gov/strbase/kinship.htm#KinshipPrograms
Advantages and Disadvantages of Likelihood

- **Advantages:**
  - Flexibility: can be extended in many ways
    - Compensate for errors in genotyping
    - Incorporate factors influencing mating success: fecundity, distance, and direction
  - Compensates for lack of exclusion power
    - Fractional paternity

- **Disadvantages**
  - Often results in ambiguous paternities
  - Difficult to determine proper cutoff for LOD score
Summary

- Direct assessment of movement is best way to measure gene flow
- Parentage analysis is powerful approach to track movements of mates retrospectively
- Paternity exclusion is straightforward to apply but may lack power and is confounded by genotyping error
- Likelihood-based approaches can be more flexible, but also provide ambiguous answers when power is lacking
Phylogenetics

- Study of the evolutionary relationships among individuals, groups, or species
- Relationships often represented as dichotomous branching tree
- Extremely common approach for detecting and displaying relationships among genotypes
- Important in evolution, systematics, and ecology (phylogeography)
Evolution

Slide adapted from Marta Riutart
What is a phylogeny?

- Homology: similarity that is the result of inheritance from a common ancestor

Slide adapted from Marta Riutart
Phylogenetic Tree Terms

- **Group, cluster, clade**
- **Terminal branches**
- **Node**
- **Interior branches**
- **Leaves, Operational Taxonomic Units (OTUs)**

Slide adapted from Marta Riutart
Tree Topology

(Bacteria1, (Bacteria2, Bacteria3), (Eukaryote1, ((Eukaryote2, Eukaryote3), Eukaryote4)))

Slide adapted from Marta Riutart
Are these trees different?

How about these?

http://helix.biology.mcmaster.ca
Rooted versus Unrooted Trees

Unrooted tree

Rooted by outgroup

Monophyletic group

Monophyletic group

Slide adapted from Marta Riutart
Rooting with D as outgroup

Slide adapted from Marta Riutart
Now with C as outgroup
Which of these four trees is different?
**UPGMA Method**

- Use all pairwise comparisons to make dendrogram

- **UPGMA:** Unweighted Pairwise Groups Method using Arithmetic Means

- Hierarchically link most closely related individuals

The pairwise evolutionary distances are given by the following distance matrix:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Also see lab 12