The Genetics of Small Populations

Random Changes in Gene frequencies
f statistics
Markov Chains
More quantitative genetics
A bit about my research
Inbreeding depression
The Four Forces of Evolution

There are EXACTLY four forces of evolution:

- Migration
- Mutation
- Genetic Drift
- Selection

The Four Horsemen
Random Change In Gene Frequency

• Fisher’s quantitative genetics uses the “infinitesimal” model.
• H-W-C equilibrium has no change in gene frequencies
• A key assumption for both models -- very large population sizes.
• Endangered species emphatically do not have “very large” population sizes!

What are the effects of small population sizes?
Random Genetic Drift

Random Changes in Gene Frequency
Genetic Drift

N = 4

N = 40

N = 400
Heterozygosity

Which gene will be fixed not predictable.

(actually the probability of fixation on an allele is equal to its gene frequency)

What is predictable is heterozygosity:

Heterozygosity = the proportion of individuals that are heterozygous, i.e., have different alleles at the locus of interest.
Heterozygosity

N = 4

N = 40

N = 400
Metapopulation: A population of populations
Modeling Drift

There are two ways of modeling drift

These are

Coancestries (f statistics)

And

“Markov chains”

We will start by looking at f statistics
# Inbreeding Coefficients

<table>
<thead>
<tr>
<th>Genotype</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>$p^2$</td>
<td>$2pq$</td>
<td>$q^2$</td>
</tr>
<tr>
<td>Random Mating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency inbred</td>
<td>$p$</td>
<td>$0$</td>
<td>$q$</td>
</tr>
<tr>
<td>Frequency partially inbred</td>
<td>$(1-f)p^2 + fp$</td>
<td>$(1-f)2pq$</td>
<td>$(1-f)q^2 + fq$</td>
</tr>
</tbody>
</table>

**Note:** \[ p = P(AA) + \frac{1}{2} P(Aa) = (1 - f)p^2 + fp + \frac{1}{2}(1 - f)2pq = p \]

QuickTime™ and a GIF decompressor are needed to see this picture.

Sneetches are a great example of assortative mating.

Sylvester McMonkey McBean
Some Definitions

Identity by State: Two alleles are identical by state if they have the same chemical makeup.
- A and A are identical by state.
- A and a are not identical by state.

Identity by descent: Two alleles are identical by descent if they share a common ancestor at some point in the past.
- These are IBD.
- These are not IBD (unless they are related further back).

Identity by descent implies identity by state. Opposite is not true.
We can define $f$ (in a metapopulation) to be the probability that the two alleles that combine to form an individual are identical by descent.

With random mating $f$ is the probability that two alleles in the same deme are identical by descent.

Wright defined it as the correlation between gametes combining to form an individual.

My definition is the same for genetic drift but not for assortative mating!
Changes in $f$

Reproduction can be thought of as sampling from the gene pool with replacement. In small populations there is a reasonable chance that the same allele will be sampled more than once. When this happens $f$ increases.

Probability that two alleles are both sampled from the same parental allele = $1/2N$. Probability they are IBD = 1

Probability that two alleles are sampled from different parental alleles = $1 - 1/2N$. Probability they are IBD = $f$

$$f_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)f_t$$
f “accumulates”

In small populations it increases, in large populations it does not decrease.

Think of an outbred population as an empty graduate cylinder.

The empty space is “heterozygosity” (=1-f)
Inbreeding continues.

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

<table>
<thead>
<tr>
<th>Gen</th>
<th>(f_t)</th>
<th>(N)</th>
<th>(f_{t+1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>infinity</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>infinity</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>2</td>
<td>0.4375</td>
</tr>
</tbody>
</table>

Inbreeding decreases heterozygosity and increases homozygosity. It “accumulates” in that large population size does not decrease.

Mutation and migration DO cause \(f\) to decrease.
Doing the Calculations

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

---

<table>
<thead>
<tr>
<th>Gen</th>
<th>(f_t)</th>
<th>(N)</th>
<th>(f_{t+1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>infinity</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>infinity</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>2</td>
<td>0.4375</td>
</tr>
</tbody>
</table>
Changes in Variance

The bottom line: In an additive system variance within demes declines, variance among demes increases, and overall there is a general increase in the variance in the metapopulation.

\[ \text{Var}(w/in) = (1 - f)V_A \]
\[ \text{Var}(bet) = 2fV_A \]
\[ \text{Var}(total) = \text{Var}(w/in) + \text{Var}(bet) = (1 + f)V_A \]
Conservation Biology

With endangered species we don’t have a “metapopulation”. We have one or a few demes.

The increase in variance among demes is of little consequence.

The loss of genetic diversity is potentially of great consequence.
The idea of a Markov chain: have a vector of population types, for example the number of A (vs a) alleles.

Example, a population of 2 individuals can have 0, 1, 2, 3 or 4 A alleles. Thus, the distribution of populations of two individuals can be described by a 4 element vector

\[ X = \begin{bmatrix}
  P(A = 4, a = 0) & 0 \\
  P(A = 3, a = 1) & 0 \\
  P(A = 2, a = 2) & 1 \\
  P(A = 1, a = 3) & 0 \\
  P(A = 0, a = 4) & 0
\end{bmatrix} \]

This is an example of a “metapopulation” in which all demes have 2 A alleles and 2 a alleles.
**Population type vectors**

\[
X = \begin{pmatrix}
P( A = 4, a = 0 ) & 0 & 0.2 & 0.5 \\
P( A = 3, a = 1 ) & 0 & 0.2 & 0 \\
P( A = 2, a = 2 ) & 1 & 0.2 & 0 \\
P( A = 1, a = 3 ) & 0 & 0.2 & 0 \\
P( A = 0, a = 4 ) & 0 & 0.2 & 0.5 \\
\end{pmatrix}
\]

The vector must add up to 1.
Transition matrices

The heart of the Markov chain model is the Transition matrix.
The transition matrix uses linear algebra to convert the population vector of one generation into the population vector of the next generation.

(for those who know linear algebra)

\[ X' = XT \]
for those who know probability, the elements of the matrix are simply binomial probabilities

\[
T_{ij} = \binom{2N}{j} \left( \frac{i}{2N} \right)^j \left( 1 - \frac{i}{2N} \right)^{2N-j}
\]
In English

\[ T_{ij} = \binom{2N}{j} \left( \frac{i}{2N} \right)^j \left( 1 - \frac{i}{2N} \right)^{2N-j} \]

\( T_{ij} \) = the probability of a population with \( i \) copies the A allele ending up with \( j \) copies in the next generation.

\( \binom{2N}{j} \) = the number of ways to get \( j \) As out of \( 2N \) draws

\( \left( \frac{i}{2N} \right)^j \) = the probability of getting \( j \) A alleles

\( \left( 1 - \frac{i}{2N} \right)^{2N-j} \) = the probability of getting \( 2N-j \) a alleles
More on the Transition matrix

\[ T = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0.316 & 0.422 & 0.211 & 0.047 & 0.004 \\
0.062 & 0.25 & 0.375 & 0.25 & 0.062 \\
0.004 & 0.47 & 0.211 & 0.422 & 0.316 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix} \]

• All rows add to 1. (you will have a gene frequency next generation)
• Rows 0 and 4 are “absorbing boundaries”
• Once a population is fixed for A or a in this model it will stay there for ever.
• All population will eventually fix for A or a
Actually calculating this stuff

The way we would actually do the calculation is:

\[
X' = X^T
\]

\[
\begin{bmatrix}
0.062 & 0 & 1 & 0 & 0 & 0 & 0 \\
0.25 & 0 & 0.316 & 0.422 & 0.211 & 0.047 & 0.004 \\
0.375 & 1 & 0.062 & 0.25 & 0.375 & 0.25 & 0.062 \\
0.25 & 0 & 0.004 & 0.47 & 0.211 & 0.422 & 0.316 \\
0.062 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
0.2764 & \cdot 1 + 0.316 \cdot 0.2 + 0.062 \cdot 0.2 + 0.004 \cdot 0.2 + 0 \\
0.1438 & \cdot 0 + 0.422 \cdot 0.2 + 0.25 \cdot 0.2 + 0.047 \cdot 0.2 + 0 \\
0.1595 & \cdot 0 + 0.211 \cdot 0.2 + 0.375 \cdot 0.2 + 0.211 \cdot 0.2 + 0 \\
0.1438 & \cdot 0 + 0.047 \cdot 0.2 + 0.25 \cdot 0.2 + 0.422 \cdot 0.2 + 0 \\
0.2764 & \cdot 0 + 0.004 \cdot 0.2 + 0.062 \cdot 0.2 + 0.316 \cdot 0.2 + 0 \\
\end{bmatrix}
\]
Markov Chain model drift

Prediction of the Wright-Fisher model for N=16 and p=0.5

From Hartl & Clark
A note on my own research

The model I showed you does not take into account gene interactions. When there is epistasis (interactions among loci) funny things can happen. One example, is that the additive genetic variance sometimes increases after a founder event!

Goodnight 1987 (Evolution 41: 80-91)
Goodnight 1988 (Evolution 42:441-454)
More on my research

It turns out this increase in VA is associated in a shift in what genes do.

After a bottleneck the VA may increase, but along with that comes a change in evolutionary trajectory.
Inbreeding Depression

From a conservation perspective the big issue is inbreeding depression

Decline in yield in corn as a function of inbreeding

Decline in litter size in swine as a function of inbreeding of the mothers and the offspring
Inbreeding in nature

These vertebrae were collected from an Isle Royale wolf whose carcass was collected in 2001. The gross asymmetry of both vertebrae represent developmental defects. These deformities may or may not be a consequence of inbreeding. It is also unknown how such a deformity might have affected the fitness and health of this wolf. It is also unknown if the observation of these deformities should be interpreted as a sign that other important genetic defects are present but undetectable.
Inbreeding in Black Grouse

In domesticated and captive populations it is known that mating between relatives may lead to reduced fitness in offspring (termed inbreeding depression). The effects of inbreeding have only rarely been demonstrated in the wild, and then only on isolated island populations. We examined the association between inbreeding and fitness in a continuous mainland black grouse population using DNA markers. We found that inbred males have reduced reproductive and competitive fitness relative to more outbred males. Our results indicate that inbreeding depression may be a general feature of natural populations.
Fireweed (*Chamerion angustifolium*)

Inbred (left) and Outbred (right) Chamerion
My Own research

*Tribolium* flour beetles.

When replicates of the same strain inbreed the outcome is highly variable.

Most lines are bad, some lines are good.

Inbreeding depression is trait specific.
Effective Population Size

Observed or Apparent population size (N): the number of individuals in the population.

Effective population size ($N_e$): The population size that would give the observed genetic behavior.

The effective population size is typically less than the observe population size

$$\frac{N_e}{N} < 1$$
Effective Population Size

There are several kinds of effective population size.

Inbreeding effective size:

Solve the inbreeding equation for $N$ given $f_t$ and $f_{t+1}$

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

Also:

Variance Effective Size
Eigen value effective size
Example: Separate Sexes

$N_m = \text{number of males}$

$N_f = \text{number of females}$

$$N_e = \frac{4N_m N_f}{(N_m + N_f)}$$

$N_m = N_f = N/2$

$$N_e = \frac{4 \frac{N}{2} \frac{N}{2}}{\left( \frac{N}{2} + \frac{N}{2} \right)} = N$$

$N_m = 1, \ N_f = \text{infinity ($\infty$)}$

$$N_e = \frac{4(1)(\infty)}{(1 + \infty)} = \frac{4(\infty)}{(\infty)} = 4$$
Genetic Drift and Molecular Evolution

Basic idea:

Mutations occur constantly over time.

• Per base-pair mutation rate is very low \(10^{-11}\)
• But there are 3.2 billion base pairs!
• This works out to about one mutation per individual every other generation
What can be predicted?

Where mutations occur cannot be predicted!
(ask me about the infinite alleles model)

The degree of divergence between populations CAN be predicted
Synonymous and non-synonymous mutations