**Finite Population Size: Genetic Drift**

READING: Nielsen & Slatkin pp. 21-27

– Will now consider in detail the effects of relaxing the assumption of infinite-population size.

– Start with an extreme case: a population of size \( N = 1 \) (an annual, self-fertilizing diploid plant).

• The sequence of events shown at right *could* occur at a particular locus:

• Notice:

  (1) Allele copies in individuals from generation 2 on are both descended from the same ancestral allele, \( c_1 \) (i.e., they are IBD)

  (2) If \( c_1 \) were an \( A \) allele, and \( c_2 \) an \( a \) allele, then the frequency of \( A \) changes from 1/2 to 1.

• Will see that these features are true of *any* finite sized population:

  (1) The level of inbreeding (homozygosity) increases.

  – eventually, all alleles will have descended from a single copy in an ancestor.

  (2) Allele frequencies will change due to randomness of meiosis.

  – eventually, the entire population will be homozygous.

  – This process of evolutionary change is called “**random genetic drift**.”

• Inbreeding and random genetic drift are two important consequences of finite population size.

  – We already discussed another when considering mutation.

– To study consequences in more detail, it will help to study the following thought experiment:

• Consider a hermaphroditic population of size \( N \) with \( 2N \) gene copies at a locus:

  

• Each individual contributes a large (but equal) number of eggs and sperm to a gamete pool.
• $N$ offspring are formed by drawing 1 egg and 1 sperm from pool at random.

• **NOTE:** Since $2N$ different allele copies can contribute to the gamete pool, the probability that a particular gene copy is drawn is $1/2N$.
  – Given that, the probability that the same allele copy is chosen again is still $1/2N$ due to the large & equal number of gametes shed by each individual.

**• Inbreeding Due to Finite Population Size**

– Consider how the inbreeding coefficient, $f_t$, changes in the population from generation $t-1$ to generation $t$.

– **Fact:** Because each generation is formed by random mating between all $N$ individuals (including selfing), the inbreeding and kinship coefficients are the identical.

– Each offspring is formed by randomly choosing 2 alleles from the parent population, so:
  
  (a) with probability $1/2N$, the same allele copy is chosen twice
    • since the same allele is being copied, the inbreeding coefficient = 1.

  (b) with probability, $1 - 1/2N$, two different parental genes are chosen
    • these genes are IBD with probability = $f_{t-1}$.

– Putting these together: $f_t = (1/2N) \cdot 1 + (1 - 1/2N)f_{t-1}$

– If $f_0 = 0$, what is $f_t$?

  • Consider $h_t = 1 - f_t = \text{Prob. of non-identity of alleles}$
  • Then $h_t = (1/2N) \cdot 0 + (1 - 1/2N)h_{t-1} = (1 - 1/2N)h_{t-1}$.
  • If $h_0 = 1$, then $h_1 = (1 - 1/2N), h_2 = (1 - 1/2N)^2, ..., h_t = (1 - 1/2N)^t$ or
    
    $f_t = 1 - h_t = 1 - \left(1 - \frac{1}{2N}\right)^t \rightarrow 1$ as $t \rightarrow \infty$.

    • i.e., Alleles at each locus will eventually be IBD with probability 1.

• The rate of approach to complete inbreeding ($f = 1$) is roughly inversely proportional to population size.

  – E.g., for 50% of the population to become inbred, it takes $\approx 14,400$ generations for populations of size $N = 10,000$, and $\approx 138$ generations for a population of size $N = 100$. 

II-11
• Genetic Drift Due to Finite Population Size

– Two views of genetic drift:

(a) Within a single population.
   • random changes in allele frequencies occur until $p = 0$ or 1 is reached; no further change occurs after that.

(b) Across replicate populations.
   • Replicate population allele frequencies diverge through time.

– Relation between the two views:

   • overall statistical properties across replicate populations are interpreted as probabilities of particular outcomes within a single population, and vice versa.

• The above idealized model was used by Wright and Fisher to study drift.

– Will refer to it as the “Wright-Fisher model.”

– Specifically assume
   • Population of size $N$ with $2N$ gene copies per locus
   • Suppose $i$ of these are $A$ alleles ($p = i/2N$)

– $Q$: How many copies of $A$ will there be in the next generation?
  $A$: It depends, unless $i = 0$ or $2N$

– Better Question: What is $P_j = Pr(N_{A}^{(i+1)} = j | N_{A}^{(i)} = i)$?

   • Since each gene copy is drawn independently, this question is mathematically equivalent to the probability of getting $j$ heads in $2N$ tosses of a coin whose probability of heads in any single toss is $i/2N$.

   • These probabilities are given by the **binomial distribution**:

     \[ P_j = \binom{2N}{j} p^j (1-p)^{2N-j} \quad \text{where} \quad p = i/2N \quad \text{and} \quad \binom{2N}{j} = \frac{2N!}{j!(2N-j)!} \]

     – From an “across populations” view, imagine replicate populations each of size $N$ and with $i$ copies of the $A$ allele, then $P_{ij} = \text{fraction of all populations with } j \text{ copies of the } A \text{ allele in the next generation.}$

– Now let’s use the Wright-Fisher model with these probabilities to study some properties of genetic drift in finite populations.
What is the average frequency of $A$ over all replicate populations?

A: Binomial expectation: $E[j] = 2Np = 2N(i/2N) = i$ or, in terms of frequencies, $\bar{p}_i = p_0 = i/2N$.

- Punch Line: No Change is expected. In fact, $\bar{p}_i = p_0$.

How much do allele frequencies vary across the (initially identical) replicate pops?

A: Binomial variance: $\text{Var}(j) = 2Np_0(1 - p_0)$ so that $\text{Var}(p_i) = p_0(1 - p_0)/2N$.

- Can show that $\text{Var}(p_i) = [1 - (1 - 1/2N)^i]p_0(1 - p_0) \rightarrow p_0(1 - p_0)$ as $i \rightarrow \infty$.

- Term in brackets should remind you of $f_i$: $f_i = 1 - (1 - 1/2N)^i$.

- In fact: $f_i = \frac{\text{Var}(p_i)}{p_0(1 - p_0)} = \frac{\text{Var}(p_i)}{\bar{p}_i(1 - \bar{p}_i)}$.

- This suggests way to estimate $f$ in an extent population.

- Remark: $f_i$ above is exactly what we found for the Wahlund Effect!??

Three Quantitative Conclusions:

1. **PROBABILITY OF FIXATION:**

   Q: If Freq($A$) = $p$ initially, what is the probability $A$ will become fixed or lost?

   - Answer 1 (replicate populations) Know:
     - All populations will eventually become fixed (i.e., $p_\infty = 0$ or $p_\infty = 1$).
     - Since the average frequency of $A$ never changes, $p$ populations must be fixed for $A$ and $(1 - p)$ will have lost $A$.
     - Probability $A$ is fixed = $p$, lost = $1 - p$.

   - Answer 2
     - In any one population, all alleles will eventually be descended from a single gene copy.
     - The chance that the lucky gene copy is an $A$ allele is just the frequency of $A$ in the original population.
     - Probability $A$ is fixed = $p$, lost = $1 - p$.

   - Note: This conclusion is independent of the population size!

2. **DECLINE IN HETEROZYGOSITY**
**Q**: What happens to the average frequency of heterozygotes?

- Let $H_i = 2p_i(1-p_i)$
- Can show $E(H_{+i}) = (1-1/2N)H_i$

- Variation is lost, but very slowly if $N$ is large.
  - e.g., if $N = 10^6$, 0.00005% of current heterozygosity is lost per generation.
  - Mendelian inheritance is thus a very powerful force for maintaining genetic variation in "large" populations (Flip side: drift is weak force in depleting genetic variation in large populations).

- Decline in expected heterozygosity does not imply heterozygote deficiencies within replicate subpopulations (as with the Wahlund effect).
  - Randomly mating subpopulations are in approximate H-W proportions.
  - The overall decline in heterozygosity is due to those subpopulations that are becoming fixed for different alleles.

(3) **TIME TO FIXATION**

**Q**: How many generations will it take for drift to cause fixation of either $A$ or $a$?

- On average, it takes $\bar{t}(p) = -4[(1-p)\ln(1-p) + p\ln p]N$ generations.

- Note that $\bar{t}(p)$ depends on $p$ and $N$
  - $\bar{t}(p) \propto N$
  - e.g., if $p = 0.5$ initially, $\bar{t}(0.5) \approx 2.7N$ generations.
  - This may be a long time for large populations.

**Population Bottlenecks**

- During population crashes or colonization events, a population may experience short periods with low numbers.
  - Numerous biologists have emphasized the importance of such "founder-flush" events in evolution.

- From a population genetics standpoint want to ask: What are the effects of drift during "population bottlenecks".
• A: Depends on
  (a) how small a population becomes.
  (b) how long it remains small.

  – Will examine the issue from two perspectives.

  (1) Effect of bottlenecks on heterozygosity

  • Consider a population bottleneck of 1 generation to $N = 2$.
    – Assume the population recovers to large size in generation 2.
  • Know that $E(H_{t+1}) = (1 - 1/2N)H_t$ or $\frac{E(H_{t+1} - H_t)}{H_t} = -1/2N$

    – In this case, only 25% of the heterozygosity is expected to be lost

  • Conclude: Appreciable amounts of heterozygosity will be lost due to drift only
    if population is small for an appreciable amount of time.

  (2) Effect of bottleneck on the number of alleles

  • Expect common alleles to persist, rare ones to be lost

  • Probability that an allele of frequency $p$ is lost during a 1-generation bottleneck
    $= p^2N$.

  • Consider the following probabilities that an allele with frequency $p$ will be lost
    during a 1-generation bottleneck of size $N$:

    | $p$  | 2   | 10  | 100 | 10,000  |
    |-----|-----|-----|-----|---------|
    | 0.5 | 0.06| $9.5 \times 10^{-7}$ | $6.2 \times 10^{-6}$ | $<10^{-999}$ |
    | 0.1 | 0.66| 0.12 | $7.1 \times 10^{-10}$ | $7.1 \times 10^{-916}$ |
    | 0.01| 0.96| 0.82 | 0.13 | $5.1 \times 10^{-88}$ |
    | 0.001| 0.9996| 0.998 | 0.98 | 0.14 |

  • Notice that rare alleles are likely to be lost, however, their loss has little effect
    on heterozygosity.

  • The time needed to recover previous heterozygosity and # of alleles depends on
    what mechanism restores variation.

    – E.g., with mutation this would take a long time to accomplish.

  • Conclude
1) Common alleles are unlikely to be lost during a bottleneck

2) Rare alleles are highly prone to being lost.

– Implications:

• If evolution relies mainly on common alleles, a few generations of small population size won’t have much effect on one population’s long-term adaptive potential.

• If, in contrast, evolution relies on rare alleles, then bottlenecks erode the ability of populations to adapt.