Overview

- Finishing up from Wednesday:
  - Haplodiploids!
  - Expected homozygosity as a function of N_e
  - Practice problem: Florida mice
- Variation in population size:
  - Cycling populations
  - Exponential growth or shrinkage
  - Bottlenecks

From the one-minute responses

- Practical examples and practice problems are good
- What are the parameters in PopG?
- Is there an equivalent coalescent simulator to play with?
  - I haven't used them, but check out the two programs at www.coalescent.dk, especially the "Hudson Animator"

N_e of a haplodiploid

- Cannot assume that half the genes came from male parents and half from females, because male haplodiploids *have no male parents*
- My reasoning:
  - Total gene copies are $N_m + 2N_f$
  - Chance a gene came from a male parent is $\frac{N_f}{N_m+2N_f}$
  - Chance a gene came from a female parent is $\frac{N_m+N_f}{N_m+2N_f}$
  - Square these to get the chance that two random genes came from a male, or a female
  - Given they came from a male, coalescence chance is $\frac{1}{N_m}$
  - Given they came from a female, coalescence chance is $\frac{1}{2N_f}$
  - So!
    $\frac{1}{2N_e} = (\frac{N_f}{N_m+2Nf})^2 \times \frac{1}{N_m} + (\frac{N_m+N_f}{N_m+2N_f})^2 \times \frac{1}{2N_f}$

Haploidiploid disclaimer

- Sewall Wright also tackled this problem (for the mammalian X chromosome)
- His solution was:

  $N_e = \frac{9N_fN_m}{2N_f+4N_m}$
- This gets the same answer when $N_m = N_f$ but not otherwise
- I don't know which is correct
- Possibly both are correct for different definitions of N_e
- (I got Joe Felsenstein to try this for an hour or so, without a numerical result....)

N_e of a haplodiploid

- $\frac{1}{2N_e} = (\frac{N_f}{N_m+2Nf})^2 \times \frac{1}{N_m} + (\frac{N_m+N_f}{N_m+2N_f})^2 \times \frac{1}{2N_f}$

- I can't reduce this, but here are some results:

Males	Females	N_e
199	1	1.01
150	50	75.00
100	100	150.00
50	150	105.00
1	199	2.00

- The equal-sexes answer is just what we would have predicted naively; there are 3/4 as many gene copies, so N_e is 3/4 what it would be in a diploid

- A shortage of females reduces N_e more than a shortage of males

Fraction of homozygotes

- Counting alleles is not a good way to quantify variation
  - Too sensitive to very rare alleles
- Measure variation as proportion of homozygotes–the fewer homozygotes, the more variation
  - Call the proportion of homozygotes F
  - With two equally frequent alleles, $F = 0.5$

Fraction of homozygotes

- In cases with mutation and drift, an approximate formula is: $F \approx \frac{1}{1+4N_e\mu}$
- This approximation assumes that every mutation is to a new allele. It is quite accurate in practice even when that's not true, as long as there are a decent number of different alleles possible.

Fraction of homozygotes

$F \approx \frac{1}{1+4N_e\mu}$

Intuitive results of this equation:

- If the population is large, there will be fewer homozygotes (more diversity)
- If the mutation rate is large, there will be fewer homozygotes (more diversity)

(Always ask yourself–does this equation predict results that are in the right general direction?)

Fraction of homozygotes–Practice problem

(Fictional problem inspired by real data of Potts et al.)

$F \approx \frac{1}{1+4N_e\mu}$

- We measure heterozygosity at one gene in the mouse MHC as 92%
- (Population: restaurant mice in Miami)
- Mutation rate (based on rat/mouse comparison) is around average for rodents: 10^{-6} per gene per generation
- How many mice does this imply, if the MHC were neutral?
- (You'll actually calculate N_e–that's okay)

Fraction of homozygotes–Practice problem

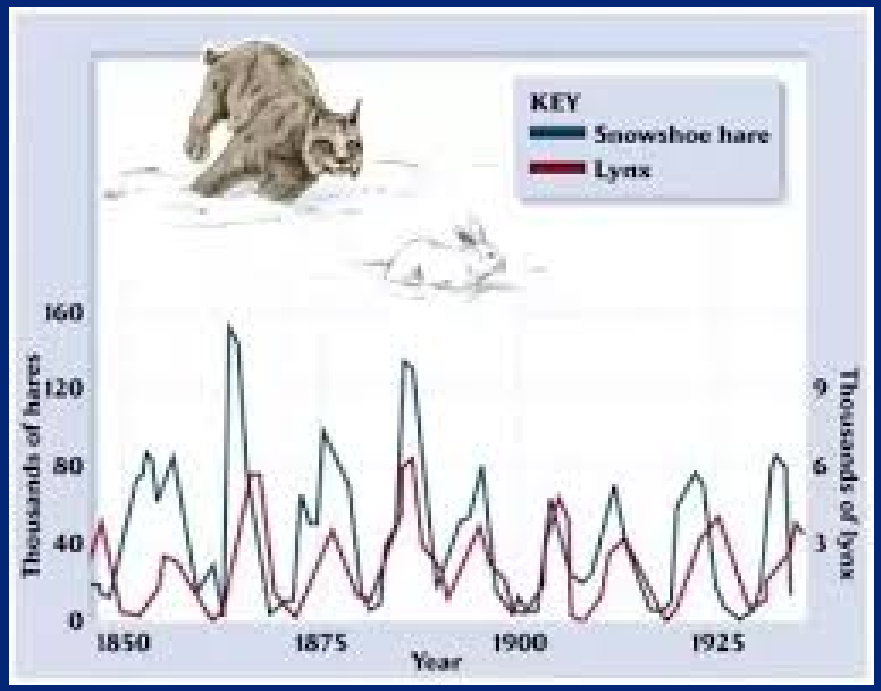
(Fictional problem inspired by real data of Potts et al.)

- $F \approx \frac{1}{1+4N\mu}$
- $0.08 \approx \frac{1}{1+4Nx10^{-6}}$
- $N = 2,875,000$ mice
- That's probably too many mice. What might explain this?

Summary

- Wright-Fisher model gives simple predictions for many aspects of the drift process:
  - Chance for a mutation to fix
  - Time it takes to fix
  - Diversity within a population
  - Divergence between populations
- These can often be adapted to a non-Wright-Fisher situation via the *effective population size* N_e

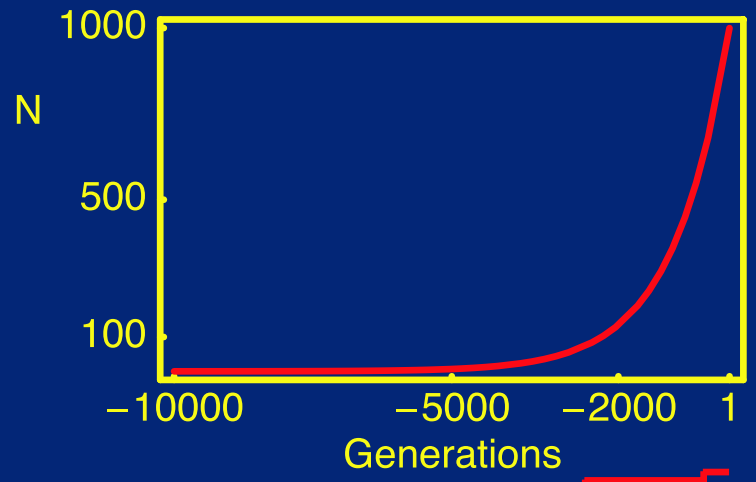
Cycling populations

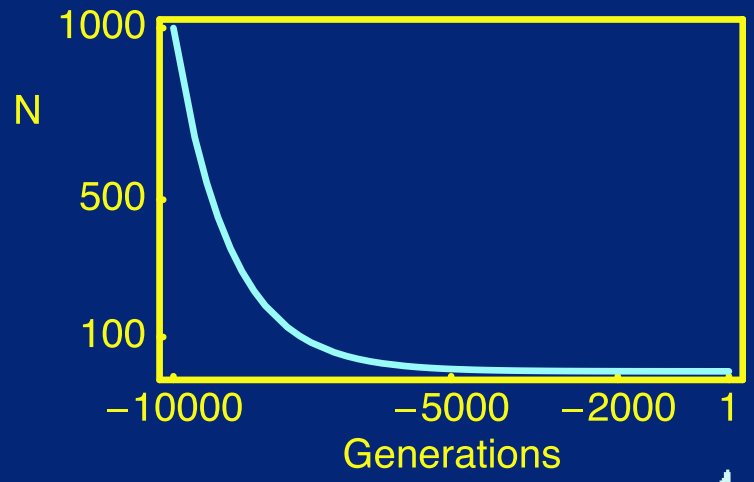


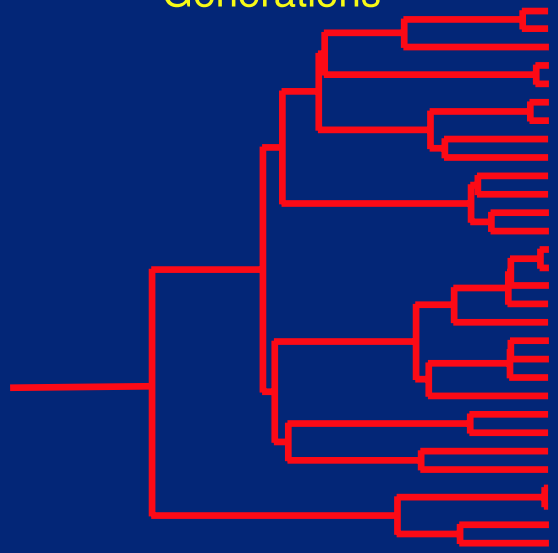
N_e in a cycling population

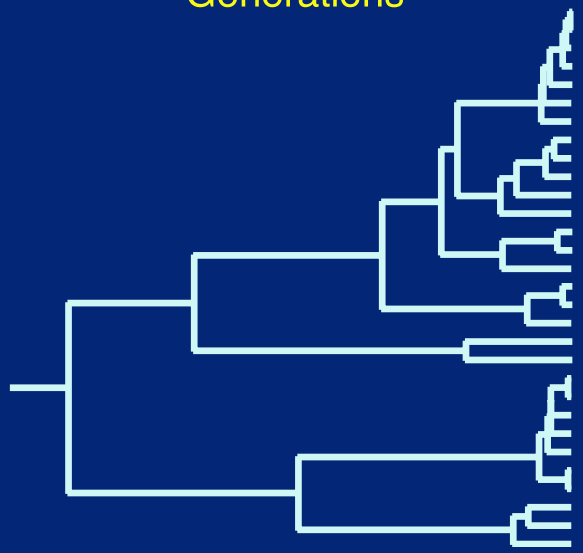
- Two generations:
  $\frac{1}{2N_e} = \frac{\frac{1}{2N_{t1}} + \frac{1}{2N_{t2}}}{2}$
- g generations:
  $\frac{1}{2N_e} = \frac{\sum_{t=1}^{g} \frac{1}{2N_t}}{g}$
- Expressed in terms of N_e, this is a $harmonic\ mean$
- It is strongly influenced by the lowest values: a cycling population has drift rates close to its minimum size
- Variability is lost when the organism is rare and not quickly regained when it is common

Exponential growth and shrinkage









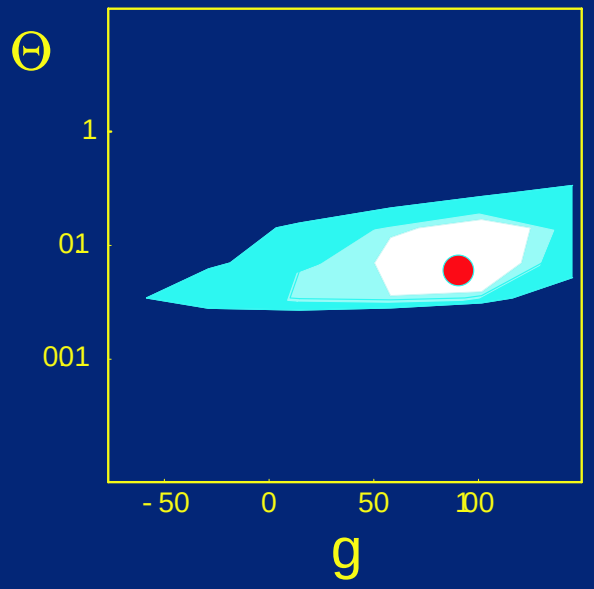
Exponential growth and shrinkage

- No general formula for N_e is possible as the drift rate keeps changing
- Over a fixed time period, can use the same logic as the cycling population
- However, shape of the coalescent is unlike that for any constant size

Estimating growth/shrinkage

- The difference in tree shape can be exploited to infer growth/shrinkage
- Weak with one locus

Water frog data: easier to estimate Θ than g



Data from a single locus were used to infer Θ and the exponential growth rate g based on the coalescent. Colors indicate 70%, 90%, 95% confidence intervals; dot is the maximum likelihood estimate. Note that the confidence interval for g includes both positive and negative g, and that it should continue well off the graph to the right.

Bottlenecks

Generations at		Approx
N=10	N=1000	N_e
1	99	503
5	95	168
10	90	92
25	75	39
50	50	20
75	25	13
90	10	11
99	1	10

Redrawn from Felsenstein textbook p. 260

Bottlenecks

- The same harmonic mean approximation from cycling populations works here
- Counter-intuitive conclusions:
  - Duration of the bottleneck makes a HUGE difference
  - Timing of the bottleneck, within the interval we're considering, does not matter

Discussion

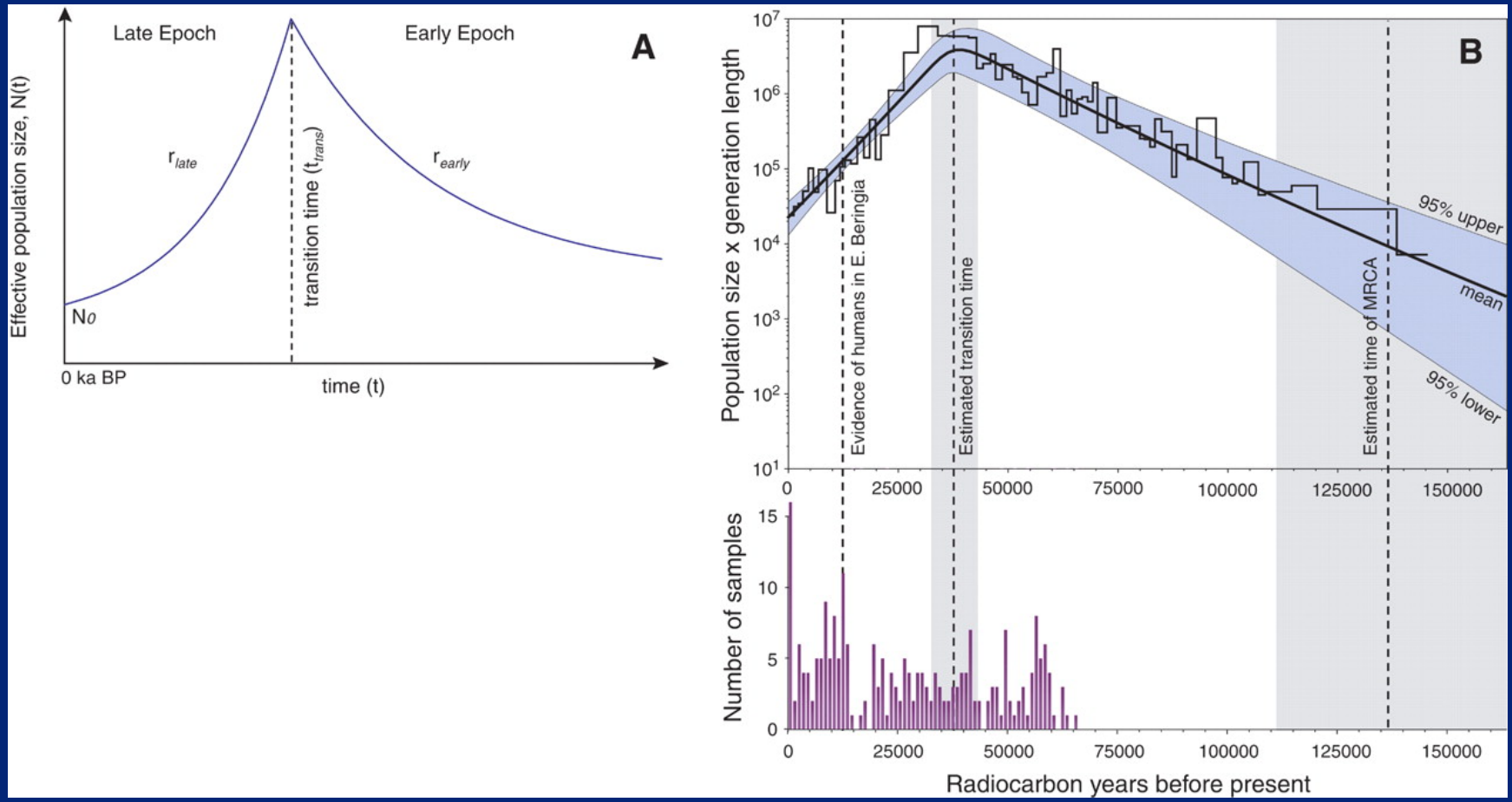
A wild population has just crashed to a low size. What does this math imply about recovery efforts?

- "The damage is done. Let's focus on making sure it doesn't get smaller."
- "We have to get the numbers back up, and the sooner the better–this is an emergency."

Inference through a bottleneck

- Bottlenecks produce a burst of coalescence in a short time
- Hard to tell this from exponential growth
- Seldom possible to look back past most recent bottleneck:
  - Many or most lineages coalesce
  - Not many lineages dating from prior to bottleneck == little information
  - Ancient DNA helps a LOT here

Beringian steppe bison



Population size of steppe bison inferred from fossil DNA. Note that time increases to the left. Shapiro et al. (2004) Science 306, p. 1561-1565

Summary

- Population diversity is measured as proportion of homozygotes (more homozygotes $==$ less diversity)
- This depends on $\theta = 4N_e\mu$
- With population growth/shrinkage:
  - N_e is harmonic mean of the different sizes
  - This is closer to the minimum size
  - Shape of the coalescent tree is distorted; coalescences pile up during periods when the population was small
  - The length of a bottleneck is very important as diversity continues to decline

Next Monday

- Models of mutation
- Mutation versus drift
  - The equilibrium state
  - Why this equilibrium is fake
- Estimation of mutation rate
- Mutation patterns as windows into mechanism

One-minute responses

- Please:
  - Tear off a slip of paper
  - Give me one comment or question on something that worked, didn't work, needs elaboration, etc.