Overview

- Models of mutation
  - 2-allele
  - K-allele
  - Infinite sites
- Equilibrium between forward and backwards mutation
  - The equilibrium equation
  - Why this equilibrium is fake
- Mutation vs. drift
- Measuring the mutation rate

From the one-minute responses

- Slow down when explaining equations!
- Define terms

Simple models of mutation

- Models with no memory:
  - 2 alleles
  - K alleles
- Models with memory:
  - Infinite sites

Watch out for "mutation rate"!

Two issues:

1. Mutation rate can be per locus or per base
   - Older theory tends to be per locus
   - Molecular studies tend to be per base
   - *Be sure you know which units are in use*

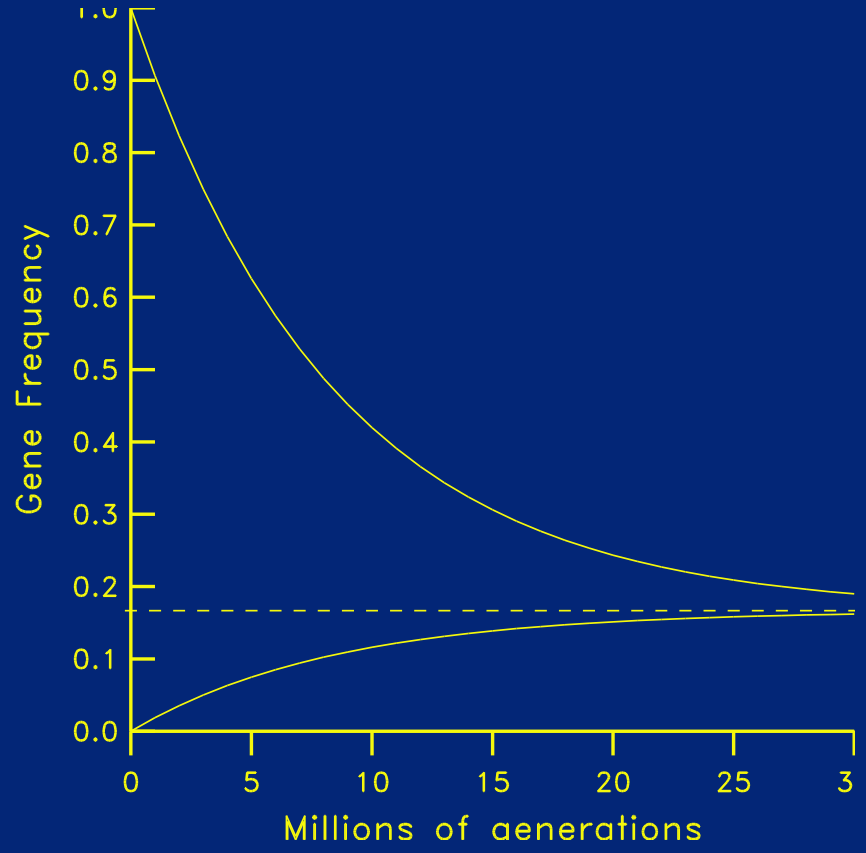
2. The rate at which mutations accumulate is not the rate at which they occur, because of selection
   - "Mutation rates are lower at the third position in a codon" – probably the rate at which mutations actually occur is identical
   - Unfortunately "mutation rate" gets (mis)used to mean accumulation rate

Simplest model: just 2 alleles

- Alleles often called A and a
- Their frequencies often called p and q
- Basis of much classical popgen theory

Forward and backward mutation

- If we think of the two alleles as "works" (A) and "doesn't work" (a)
- A mutates to a at rate μ
- a mutates to A at rate ν
- μ is generally MUCH bigger than ν
- Equilibrium at $\nu/(\mu + \nu)$



Approach of allele frequency ("gene frequency" is a common but confusing name for this) to equilibrium with $\mu = 10^{-7}$ and $\nu = 2 \times 10^{-8}$, from Felsenstein text p. 139.

Deriving the equilibrium formula

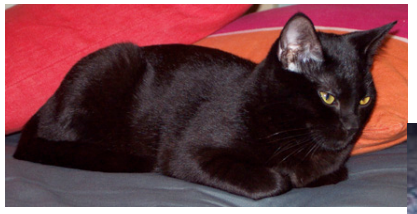
- Define p as frequency of A
- Calculate one-generation change in p as function of μ and ν:
- $p' = p(1 - \mu) + (1 - p)\nu$
- At equilibrium, $p' = p$

This equilibrium isn't real

- Assumes that all a can mutate to A with rate ν
- As a gene accumulates more and more damage, ν would get smaller
- At the real equilibrium all sequences are equally probable, which means functional genes would be vanishingly rare
- Equation is useful for the early trajectory, e.g. when a gene has just lost its function and begins to deteriorate

Sometimes two alleles just aren't enough

Different alleles with different outcomes… an example



C?

Burmese—c^bc^b or c^bc^s or c^bc

C gene codes for tyrosinase... 1st step in melanin synthesis

Siamese
c^sc^s or c^sc

cc

What are these other alleles?

How to explain the unusual phenotypes?

Going beyond 2-allele model: two directions

- K-allele model
  - Allow more than one allele
  - Any allele mutates to any other (often, at the same rate)
  - No detectable relationship among different alleles
  - (Did Burmese allele arise from Siamese allele?)

- Infinite-sites model
  - Imagine sequence of infinite length (and infinitesimal mutation rate)
  - Every mutation is to a new allele
  - Every allele carries all its ancestors' mutations
  - Can see relationships among alleles

False but useful: simplified mutation models

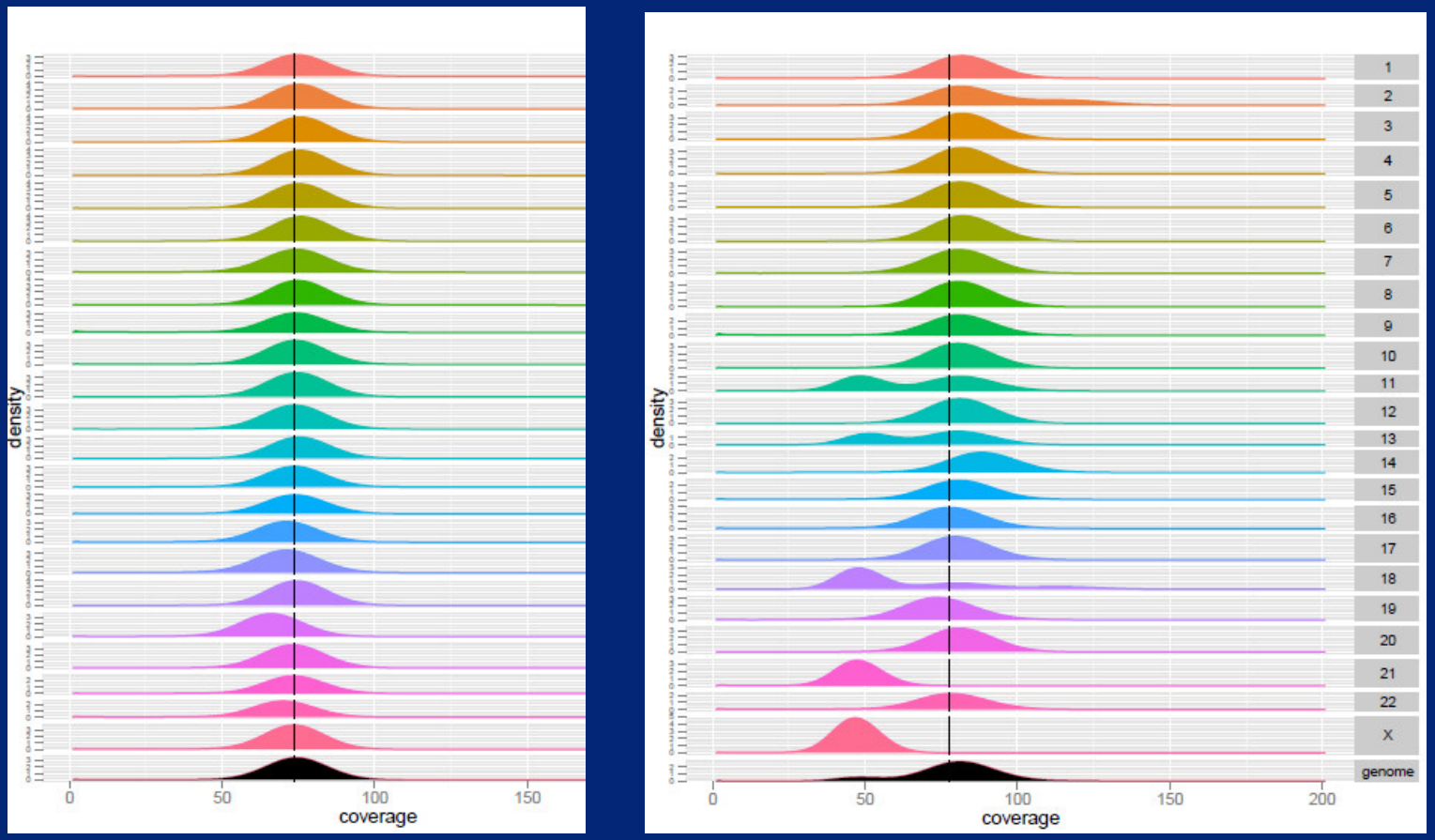
- K-allele model:
  - Asserts there are only a certain number of possible alleles, and any allele can mutate to any other allele
  - Allows back mutation and convergent mutation
  - Model used for much of classical population genetics (often K=2)
  - Useful when we mainly care about phenotypes
  - A perfect match for single nucleotide data (K=4)

False but useful: simplified mutation models

- Infinite-sites model
  - Allows alleles to have history
  - No gene really has infinite sites; model denies back mutation and convergent mutation
  - Breaks down in high-mutation organisms (HIV)
  - Extremely widely used for long DNA sequences because it $almost$ fits and is simple

An infinite-sites issue in cancer research

- If a sample is a mix of different cell lines from a tumor, can we sort them out?
- "Mutations" here are somatic (within a single patient)
- Much work on this uses infinite-sites:
  - $ONLY$ descendants of the original mutant have the mutation
  - ALL descendants of the original mutant have the mutation
  - This greatly simplifies the math, but....



Read depth of whole-genome sequencing in two samples from a pre-cancer tissue. Note that the individual is female, as shown by the first sample, but the second sample has apparently lost one copy of chromosomes 18, 21, and X as well as parts of 11 and 13.

An infinite-sites issue in cancer research

- When cells lose whole chromosomes, they must lose somatic mutations too
- Some papers have tried to argue that $important$ mutations are not lost
  - This doesn't hold up logically
- Other papers omit all regions with abnormal copy number
  - In some of our samples this omits...everything
- I am submitting an NIH grant proposal which promises to fix this

Mutation vs. drift

- With no mutation, drift eventually removes all but one allele
- With mutation, an equilibrium is possible
- General principle:
  - $4N$ generations back to population MRCA (in a diploid)
  - How much will another force accomplish in $4N$ generations?
- Variation will normally be present when $4N\mu \gg 1$

Watch out for rates!

- Was that per locus or per base pair?
- Per base pair:
  - Human $4N_e\mu$ per base pair may be around 0.0001
  - Most bases do not vary substantively in the population (are not SNPs)
- Per locus:
  - μ per locus varies a lot
  - $4N_e\mu$ for many loci is above 1 and variability is expected

Phenotypic per-locus mutation rates

Table taken from Farnsworth 1978.

E. coli	histidine auxotrophy	$2x10^{-6}$
	streptomycin sensitivity	$1x10^{-8}$
	phage T1 resistance	$2-3x10^{-8}$
Drosophila males	brown eyes	$3x10^{-5}$
	eyeless	$6x10^{-5}$
	yellow body	$1.2x10^{-4}$
Corn	colorless kernel	$2x10^{-6}$
	shrunken kernel	$1.2x10^{-6}$
Human	achondroplasia	$1x10^{-5}$
	aniridia	$2.9x10^{-6}$
	retinoblastoma	$6-7x10^{-6}$

What are these phenotypic rates?

A composite of:

- Per-base mutation rate
- Gene size
- Proportion of mutations that change the phenotype
- Mutants that survive long enough to be counted

How to estimate mutation rate?

- Compare parents to offspring or raise cells in culture
  - Very labor-intensive if the rate is low
  - Does not eliminate natural selection completely
- Measure differences between two lineages whose TMRCA is known
  - Discussion: what are hazards of this approach?

Improving the TMRCA method

- Try to find something not under natural selection:
  - Third codon positions?
  - Introns?
  - Intergenic regions?
- Incorporate uncertainty about TMRCA
- Use more than one calibration point
  - e.g. human/chimp, ape/monkey, primate/rodent

Wednesday

- Hardy-Weinberg
- Natural selection

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.