Roadmap

- Chi-square test
- Mutation:
  - What kinds of mutation can occur?
  - How common are they?
  - What is the long-term effect of mutation by itself?

Chi-square test

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

Chi-square test

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

Genotype	Observed	Expected	$(o-e)^2/e$
HH	750	640	18.9
Hh	100	320	151.3
hh	150	40	302.5
			472.7

Chi-square test

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

Genotype	Observed	Expected	$(o-e)^2/e$
HH	750	640	18.9
Hh	100	320	151.3
hh	150	40	302.5
			472.7

Table lookup:

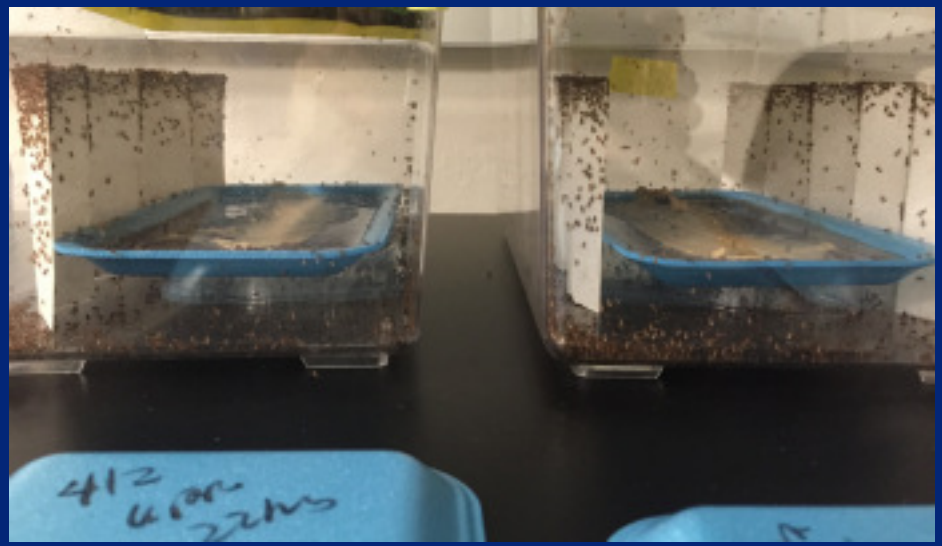
df	0.10	**0.05**	0.025	0.01	0.005
1	2.706	**3.841**	5.024	6.635	7.879
2	4.605	**5.991**	7.378	9.210	10.597
3	6.251	**7.815**	9.348	11.345	12.838

Chi-square test – degrees of freedom

- To test H-W with three categories:
  - Start with three observations
  - Use one up to get totals
  - Use one up to get allele frequencies
  - Left with 1 df

Example: Dobzhansky and Levene 1948

- Two different gene orders on chr. 3 in *D. pseudoobscura*
- Wild animals were bred in a large "population cage"
- Not in H-W: too many heterozygotes were observed
- What are some possible explanations?



Example: Dobzhansky and Levene 1948

- Females were taken from the cage and their eggs genotyped
- Eggs were in H-W for the two gene orders
- Which explanations does this rule out?

Chi-square and degrees of freedom revisited

- We mate 20 heterozygous red-eyed flies with 20 homozygous white-eyed flies (red is dominant)
- Offspring look like this:

Red Rr	66
White rr	34

- Start with 2 observations
  - Lose one for totals
  - Do NOT lose one for allele frequencies (already known)
- 1 df: χ^2=10.24, p=0.0014, significant

Chi-square and degrees of freedom revisited

- What if we had a box of random red and white eyed flies, and counted offspring?

Red $R?$	66
White rr	34

- Start with 2 observations
  - Lose one for totals
  - Lose one for allele frequencies
- Zero are left–$no\ test\ is\ possible$

Chi-square and degrees of freedom revisited

- Let's use snapdragons instead, where Rr is pink
- Random mix of red, white and pink parents gives us:

Red RR	17
Pink Rr	49
White rr	34

- Start with 3 observations
  - Lose one for totals
  - Lose one for allele frequencies
- 1 df: χ^2=0.194, p=0.9074, not significant

Resources

- Wikipedia article on Hardy-Weinberg is quite good
- Chi-square calculator at
  http://www.graphpad.com/quickcalcs/chisquared2.cfm

Mutation

- Any heritable change in the genetic material
- Provides raw material to evolution
- Without mutation, variation would be lost by:
  - natural selection
  - genetic drift (random chance)
- ...and evolution would eventually stop

Roadmap

- What kinds of mutations occur?
- What causes them?
- How frequent are they?
- What are their effects?
- What are the long-term implications of mutation on its own?
- Is mutation rate itself under natural selection? What rate would be optimal?

Point mutations

- coding changes (missense mutations)
- silent-site changes
- stop codons (nonsense mutations)
- control region changes
- changes in unused regions ("junk")

Other mutations

- Insertions and deletions
  - Frameshifts–insertions or deletions that change the reading frame
- Duplications–multiple copies of a region
- Inversions–reverse gene order
- Translocations–move region elsewhere
- Transposon insertions

Causes of Mutation

- Replication errors
- Chemical damage–can include crosslinked bases, modified bases
- Radiation damage–often single and double-strand breaks
- Transposition
- Viral insertion
- Unequal crossing-over

Mutation Rates

Table taken from Farnsworth 1978. These are rates per locus, not per site; they were estimated by observing phenotypes.

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	$6-7x10^{-6}$

Some critters are sloppy

Rates per bp per generation:

Human nuclear genome	$1x10^{-9}$
mtDNA control region	$6x10^{-6}$
HIV virus	$1x10^{-3}$

- Difference is effort into proofreading

The Standard Genetic Code

First Position (5' end)	Second Position				Third Position (3' end)
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met Start	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Silent, coding, and control mutations

- The protein sequence is changed by:
  - Almost all 1st position changes
  - All 2nd position changes
  - Relatively few 3rd position changes
- If gene is important only via its protein, mutations which don't change the protein are unimportant
- This is not always true:
  - regulatory regions
  - splice sites
  - codon bias

Mutation Rates

- Silent and coding sites generally have the same underlying mutation rate
- However, many mutations at coding sites are lost
- Jargon:
  - Mutation rate: how rapidly do mutations occur?
  - Substitution rate: how rapidly do mutations accumulate?
- Silent positions generally have equal mutation rate, but higher substitution rate
- It is common but sloppy to refer to substitution rate as "mutation rate"

Mutation without selection

- Mutation rates at most loci are asymmetrical
- Easier to break a gene than fix it
- Mutation from A to a is μ (mu)
- Mutation from a to A (back mutation) is ν (nu)
- Equilibrium reached at:

$$pA = \frac{\nu}{\nu + \mu}$$

- Usually pA is very small at equilibrium.

Mutation without selection

- When $\mu > \nu$ genes deteriorate without selection
- For organisms with "normal" mutation rates (around 10^{-9} per bp) this process is VERY slow

Mutation without selection

Suppose there are 100 sites in a gene which will destroy function if they mutate, and each mutates with a probability of 10^{-9}. Reverse mutation has to hit the same site, and has to restore the old base pair.

$\mu = 10^{-7}$

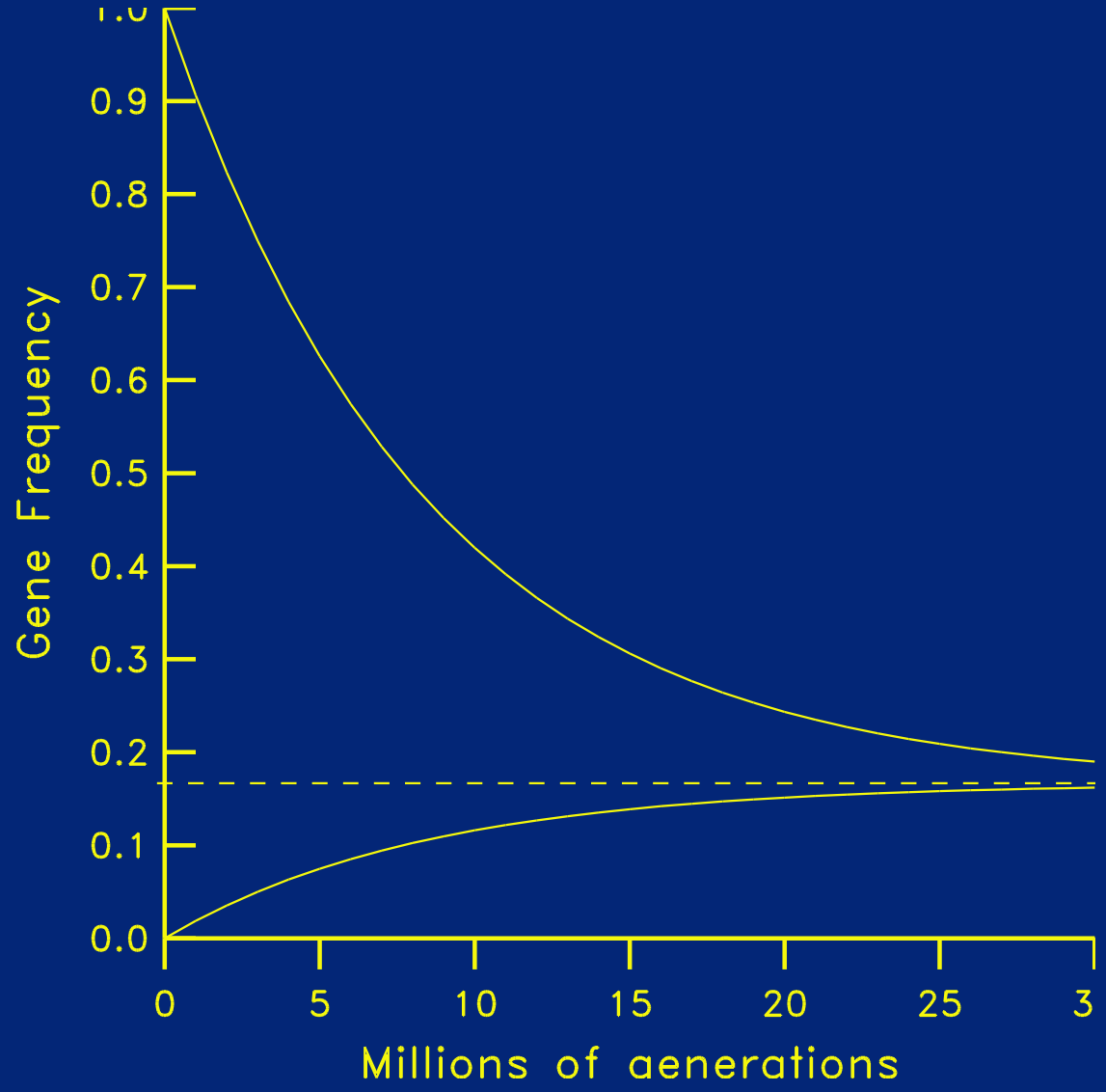
$\nu = 0.33x10^{-9}$

What happens in one generation?

Allele	A	a
Frequency before mutation	0.9	0.1
Frequency after mutation	0.89999991	0.10000009

(The effect of reverse mutation is too tiny to see!)

Mutation without selection



Genomic deterioration

- Note the millions of generations in previous slide
- If we remove selection from a genetic disease:
  - Gene will eventually deteriorate
  - 1 million human generations = 20 million years
  - Human cultures and institutions aren't stable on this time scale....
- A bigger concern is alleles that are favorable with medicine but unfavorable without it

Genomic deterioration–Practice problem

- HIV-1 virus mutation rate of 0.001 per base per generation
- Assume 100 critical bp in the env gene
- A reverse mutation has to restore original sequence
- What is pA (frequency of healthy allele) after 1 generation?
- (Ignore chance of multiple hits to the same virus)

Genomic deterioration–Practice problem

- HIV-1 virus mutation rate of 0.001 per base per generation
- Assume 100 critical bp in the env gene
- A reverse mutation has to restore original sequence
- What is pA (frequency of healthy allele) after 1 generation?
- (Ignore chance of multiple hits to the same virus)
- $\mu = 0.001 * 100 = 0.1$
- $\nu = 0.001/3 = 0.00033$
- $pA = 0.9$

Genomic deterioration–Practice problem

$$pA = \frac{\nu}{\nu + \mu}$$

- What is the equilibrium pA?
- What would happen to this gene if it were not under selection?

Genomic deterioration–Practice problem

$$pA = \frac{\nu}{\nu + \mu}$$

- At equilibrium, $pA = \frac{\nu}{\nu+\mu} = \frac{0.00033}{0.10033} = 0.0032$
- Without natural selection, mutation would rapidly destroy the HIV genome

This equilibrium is fictional

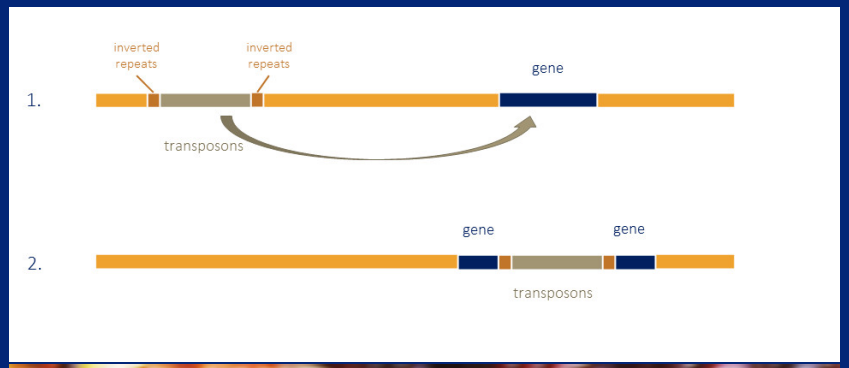
- In reality there are always more than 2 possible alleles
- Once a gene starts deteriorating it will accumulate more mutations
- The number of functional copies will eventually drop to 0
- The equilibrium approach curve is more useful in showing early steps of this process

Is mutation good or bad?

- Most mutations have no fitness effect
- Of those that do, most are bad
- Most organisms expend significant energy trying to avoid mutations (DNA proofreading, etc)
- Are organisms trying (and failing) to reach a mutation rate of zero?
- Could there be selection in favor of a non-zero rate?

Transposons as mutagens

- Transposons are genetic elements that can move around the genome
- This causes mutations:
  - Break up a coding sequence
  - Separate a gene from its control region
  - Introduce a new control region

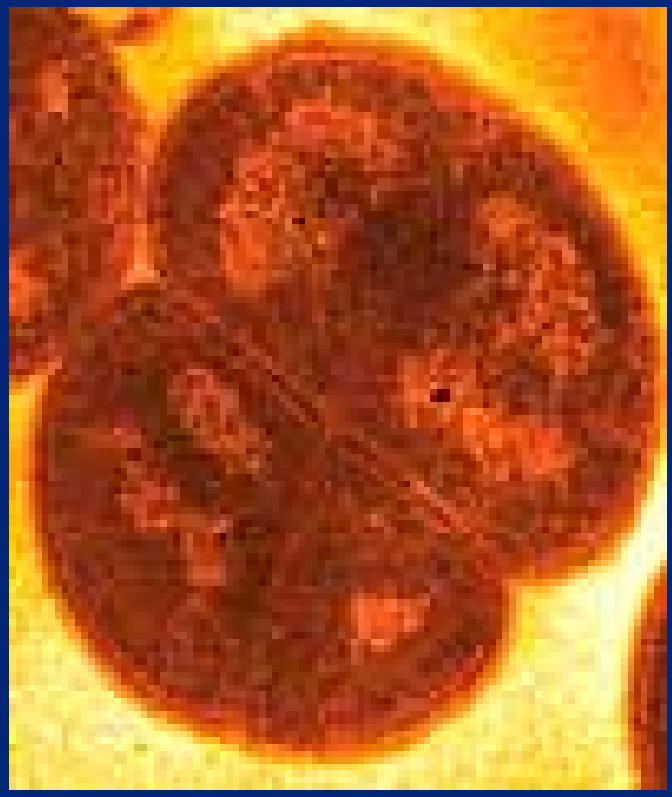


McClintock's genome shock hypothesis

- Transposition in maize increases when the plant is stressed
  - drought
  - salt
  - insects
- Transposition could be adaptive ("genome shock" theory)
  - Gives chance to fix the bad situation
- Transposition could be a symptom of illness
  - Transposons need to be kept under control
  - A sick plant can't do it

How low can mutation rate go?

- Discovered growing on irradiated meat
- Can withstand 1000x as much radiation as a human cell:
  - Chromosomes broken into about 100 pieces
  - Growth stops while chromosomes are repaired
  - Very few mutations result



Deinococcus radiodurans

Very good replication fidelity is possible

- *D. radiodurans* has 4 copies of its genome (redundant backups)
- Natural environment sunny, salty, and hot
- All three can damage DNA
- Engineered *D. radiodurans* may be useful in biodegrading radioactive chemical waste

Very good replication fidelity is possible

Presumably other cells could repair as well as *D. radiodurans*, but they don't.

- Redundant backups are expensive
- Repair machinery is expensive
- Mutations are expensive too (many are bad)
- Too-low mutation rate might inhibit adaptation
  - Hard to test this: it's a long-term effect

Mutation rates in perspective

- Human genome has $6x10^9$ bp.
- Point mutation rate around $1x10^{-9}$ per bp per generation
- Human population around 7 billion
- Every point mutation compatible with life exists somewhere
- Every human has several new point mutations

Can I join the X-Men?

- Most of these mutations are not in genes, and have little to no effect
- Many of the remaining ones are silent
- Most of the coding mutations are harmful
  - Most harmful mutations are recessive
  - Problem for your offspring, not for you....
- New beneficial mutations are rare
- Very different from Hollywood image of "mutants"

One-minute responses

- Tear off a half-sheet of paper
- Write one line about the lecture:
  - Was anything unclear?
  - Did anything work particularly well?
  - What could be better?
- Leave at the back on your way out