SGCEP BIOL 1020K
Introduction to Biology II
Spring 2012 Section 20587

Steve Thompson: steventhompson@sgc.edu
http://www.bioinfo4u.net/
Natural Selection and Variation through Mutation

Or “survival of the fittest.” But what does that really mean? Certainly not the biggest, baddest ass around. Actually to be the fittest in this context is to pass on more of your genes to the next generation than your peers, i.e. to have higher reproductive success. But how, really, does it work?
One more time — define natural selection:

* The differential reproductive success of individuals with particular genotypes (in modern lingo). Or as Darwin stated, “the preservation of favourable variations and the rejection of injurious variations.”

* However, as I pointed out before, the particular combination of traits selected are advantageous in a given situation based on history, not prediction.
Natural selection is but one mechanism involved in evolution, but it’s certainly a powerful one, particularly in “microevolution,” i.e. relatively short-term changes in allele frequencies within a population or species.

* But where does this necessary variation come from?
* Primarily from mutation. This is the random occurrence of changes in an organism’s DNA, and in sexual organisms this has to be in the germ line (egg or sperm) in order to be inherited.
Other sources of variation include . . .

* migration between populations (gene flow),

* shuffling of genes through sex (independent reassortment and recombination), and . . .

* horizontal gene transfer and hybridization.
Natural selection eliminates certain phenotypes.

- Mendel’s work although contemporaneous with Darwin’s, was not well known, so the mechanisms of how all this worked remained obscure. Now we now know that . . .
- Gene pool – entire collection of genes and their alleles in a population.
- Proportion of alleles for each gene determines characteristics for that population.
- A change in allele frequency in a population is evolution.
- Poorly adapted phenotypes are ‘weeded out.’
- Adaptive phenotype in one set of of circumstances may be a liability in others.
- Constantly changing conditions means evolution never stops.
Natural selection does not have a goal!

- It does not lead to more ‘perfect’ organisms — it can’t look ahead, nor predict what will work.
- Every genome has a limited potential.
- No gene pool contains every allele needed to confront every possible change in the environment.
- Disasters can indiscriminately wipe out the best allele combinations just by chance — extinction is the rule, not the exception.
- Some harmful genetic traits are out of natural selection’s reach (e.g. if they appear after reproductive age).
But to understand this you need — population genetics and the neutral theory

OK, this is gonna get a little hairy, but hang in there. It really does make sense, and it really helps to explain most everything natural selection (and sexual selection) doesn’t handle so well. In other words — it fills in all the background and explains what happens in populations in the ABSENCE of selection, just due to random effects. This is known as the “null” model in science.
Which brings us to “models” in science . . .

“A model is an intentional simplification of a complex situation designed to eliminate extraneous detail in order to focus attention on the essentials of the situation” (Daniel L. Hartl).
So what’s a ‘good’ model?

A good model balances fit of the data with simplicity, somewhere between under- and over-fitting — prefer the simplest model that adequately explains the data for the purpose at hand, e.g:

\[ y = 1.30 + 0.965x \]
\[ (r^2 = 0.963) \]

\[ y = -330 + 134x - 15.5x^2 + 0.816x^3 \]
\[ -0.0225x^4 + 0.00335x^5 \]
\[ -0.0000255x^6 + 0.00000000777x^7 \]
\[ (r^2 = 1.000) \]
What’s this have to do with population genetics and evolution?

- Models are used, and in particular the null model is used, as a ‘baseline’ that we can compare ‘reality’ to and attempt to explain why we see differences.
- The deterministic model we need at this point is called “Hardy-Weinberg equilibrium.”
- It is the highly unlikely situation in which allele frequencies do not change from one generation to the next (G.H. Hardy and, independently W. Weinberg, 1908).
The only way this can occur is if:

* Mutations do not occur at all,
* Mating is absolutely random,
* There’s no migration in or out of a population,
* The population is large enough (actually infinite) to avoid random changes in allele frequencies (genetic drift*), and
* Natural selection is not occurring!

Obviously this never happens.

* More on this in a just a bit.
Nonetheless . . .

* This is a great null model, as we can see how real populations deviate from it, and formulate hypotheses based on those comparisons.

* Here are the probability equations:

  * $p + q = 1$. Describes the frequency of two alleles in a population of diploid organisms, if only those two alleles exist. Then, it also holds that . . .
  * $p^2 + 2pq + q^2 = 1$, where $pq$ is the heterozygote; i.e:
    * $DD + Dd + dd = 1$

* This is calculated in successive generations to look for changes in allele frequency.
One more time . . .
with idealized ferrets . . .

\[ p = \text{frequency of } D \text{ (dominant allele)} = \text{dark fur} = 0.7 \]
\[ q = \text{frequency of } d \text{ (recessive allele)} = \text{tan fur} = 0.3 \]

<table>
<thead>
<tr>
<th>Algebraic Expression</th>
<th>What It Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ p + q = 1 ]</td>
<td>Frequency of all dominant alleles plus frequency of all recessive alleles for this gene.</td>
</tr>
<tr>
<td>[ p^2 + 2pq + q^2 = 1 ] ((DD + 2Dd + dd = 1))</td>
<td>For a particular gene, the frequencies of all the homozygous dominant individuals (p^2) plus heterozygotes (2pq) plus all homozygous recessives (q^2) add up to all of the individuals in the population.</td>
</tr>
</tbody>
</table>
Huh?

Remember:
Presumed allele frequencies –
Dark 70%
Light 30%

### Generation 1

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Dark</th>
<th>Dark</th>
<th>Tan</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Genotype frequencies</th>
<th>$DD$</th>
<th>$Dd$</th>
<th>$dd$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.49</td>
<td>0.42</td>
<td>0.09</td>
</tr>
</tbody>
</table>

| Allele frequencies | $D$: 0.7 | $d$: 0.3 |

Population mates at random

### Generation 2

| Female gametes | $Dp = 0.7$ | $dq = 0.3$ |

<table>
<thead>
<tr>
<th>Male gametes</th>
<th>$Dp = 0.7$</th>
<th>$dq = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DD$</td>
<td>$p^2 = (0.7)^2 = 0.49$</td>
<td></td>
</tr>
<tr>
<td>$Dd$</td>
<td>$pq = (0.7)(0.3) = 0.21$</td>
<td>$dq = (0.3)^2 = 0.09$</td>
</tr>
</tbody>
</table>
Here's all the numbers:

<table>
<thead>
<tr>
<th>All possible crosses</th>
<th>Genotype frequency in offspring</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
<td>$DD$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0.49 $DD$</td>
<td>0.49 $DD$</td>
<td>0.2401</td>
</tr>
<tr>
<td>0.49 $DD$</td>
<td>0.42 $Dd$</td>
<td>0.1029</td>
</tr>
<tr>
<td>0.49 $DD$</td>
<td>0.09 $dd$</td>
<td>0.0441</td>
</tr>
<tr>
<td>0.42 $Dd$</td>
<td>0.49 $DD$</td>
<td>0.1029</td>
</tr>
<tr>
<td>0.42 $Dd$</td>
<td>0.42 $Dd$</td>
<td>0.0441</td>
</tr>
<tr>
<td>0.42 $Dd$</td>
<td>0.09 $dd$</td>
<td>0.0189</td>
</tr>
<tr>
<td>0.09 $dd$</td>
<td>0.49 $DD$</td>
<td>0.0441</td>
</tr>
<tr>
<td>0.09 $dd$</td>
<td>0.42 $Dd$</td>
<td>0.0189</td>
</tr>
<tr>
<td>0.09 $dd$</td>
<td>0.09 $dd$</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

$$DD + Dd + dd = 1$$
So what? Well, when the numbers don't stack up, when allele frequencies do change, we know evolution is occurring, and we can try to explain how.

How?
- Natural Selection,
- Mutation,
- Nonrandom mating (choice!),
- Gene flow (migration), and/or . . .
- Genetic drift (randomness).
It's really not too hard and makes intuitive sense. Don't get too hung up on the mathematics—I won't.
Hardy-Weinberg we just saw. And we talked about variation through mutation, but remember...

- Mutation — provides the raw material for evolution — the variation for evolution to act upon.
- However, it is not ‘directed.’ Genes don’t ‘know’ to mutate in any particular way. Antibiotics do not create drug resistance. The mutation didn’t occur because the bacteria ‘needed’ it. It just happened (shit happens), and among all the mutations present, some happen to confer antibiotic resistance! Therefore, those ‘sweep’ the population.
- Resistance arises in response to drug exposure — the drug creates a situation in which a mutant (if it’s there before exposure to the drug) could flourish.
- In humans each germ-line genome has about 120 new mutations per generation. Plenty for evolution to occur.
Sexual selection! Choice matters, ya think . . .

- Nonrandom mating — many factors influence mate choice in many types of critters.
- The net effect is a concentration of particular alleles in particular populations.
- Variation in the ability to attract mates — may result in sexual dimorphism. Ideas of runaway sexual arms races.
- Can be a HUGE, even counterintuitive, effect.
Nice book on this topic . . .

* "It takes all the running you can do, to keep in the same place." The Red Queen in Lewis Carroll’s Alice in Wonderland Through the Looking-Glass. It’s an evolutionary arms race.

* "For an evolutionary system, continuing development is needed just in order to maintain its fitness relative to the systems it is co-evolving with." Even if that coevolution is between the sexes!

Gene flow — migration in or out of a population.

* Individuals or gametes can disperse.

Around a pond, or around the world!
Genetic drift – change in allele frequencies due to chance (i.e. more shit happens) . . .

- Changes aren’t directional, they’re merely random.
- They are due to the random sampling (no longer an infinite population) of a large number of gametes at each generation. Two huge, related impacts:
  - Founder effect — a small group of individuals forms a new population. Versus . . .
  - Bottleneck effect — many members of a population die leaving only a few survivors.
- Either way, there’s a limited set of alleles available, a rapid loss or fixation of those alleles that are present, and therefore, a subsequent loss of genetic diversity.
- Whereas, with large populations there’s more allele variability available, and more random chance for any of them to be passed on.
Trajectories of genetic drift:

* Population size matters!
* Even ‘good’ alleles can be lost;
* And ‘bad’ alleles can be fixed*; especially in a small population.

*Fixation — all members of a population have a particular allele at 100% frequency, i.e. in a diploid population everybody’s homozygous for the same allele in a particular gene.

http://www.bio.fsu.edu/~stevet/VSU/animations/Chapter13/genetic_drift.swf or http://highered.mcgraw-hill.com/sites/dl/free/0072835125/126997/animation45.html and see...

http://psych.colorado.edu/~carey/hgss/hgssapplets/evolution/geneticdrift/GeneticDrift.html
Genetic Drift: another model . . .

Basics: Wright-Fisher population model

All individuals release many gametes and new individuals for the next generation are formed randomly from these.

R.A. Fisher and Sewall Wright, independently, from 1930’s through the ’60’s.
The Wright-Fisher population model. Again, a set of overly simplistic assumptions:

* Population size N is constant through time.
* Each individual gets replaced every generation.
* The next generation is drawn randomly from a large gamete pool.
* Only genetic drift affects the allele frequencies, i.e. no selection, etc.

We can use this model to look back in time within a population.
This process is directly related to population size.

Large populations need more time to coalesce to a single lineage than small populations.

Probability of fixation for a new, neutral mutation = $1/(2N)$ and time to fixation = $\sim 4N$;
Probability of loss = $1 - (1/(2N))$ and time to loss = $2\ln(2N)$, where $N$ is the population size.
With this model, and a bunch of fancy math . . .

- We can calculate backwards and come up with all sorts of neat things like:
  - How big the ancestral population was and whether it's growing or shrinking,
  - How long ago the population started,
  - Even things like recombination and migration rate.

The expectation for the time interval $u_k$ is

$$\mathbb{E}(u_k) = \frac{4N}{k(k-1)}$$

$p(G|N) = \prod_i \exp\left(-u^k_{i,k-1} \frac{1}{4N} \frac{1}{2N}\right)$

Again, deviations from the model can be due to selection and the other factors that the model discounts, but they can be modeled as well and calculated!
This really did happen with Cheetahs, circa late 1800's...
What about “the neutral theory of evolution?”

* Motoo Kimura (1968): Most evolutionary change at the molecular level, and therefore, most variability within species, is due to random genetic drift of mutant alleles that are effectively neutral or nearly so. Many mutations are still deleterious and quickly removed from a population, it’s just that most of the others are neutral.

* Consequently most of the allelic variability (polymorphism) that we observe, at any given time point, is merely a ‘snapshot’ of alleles that are either on their way to fixation or extinction.

* Again, this is a null model, and deviations from it point to instances of positive selection!
This is easy to see in real data.

If you compare the human and mouse genome then . . .

© 2008 Nature Publishing Group, data from (Makalowski, W., et. al., 1998) and (Zheng, D., et. al., 2007)
This neutral theory claims that the overwhelming majority of evolutionary changes at the molecular level are not caused by selection acting on advantageous mutants, but by random fixation of selectively neutral or very nearly neutral mutants through the cumulative effect of sampling drift (due to finite population number) under continued input of new mutations (Kimura, 1991).

One strong (and elegant) prediction of the neutral theory is that at selectively neutral sites, the rate of substitution is equal to the rate of mutation.

Therefore, the rate of substitution per generation \( K \) is obtained simply by multiplying the number of mutations that occur at each generation by their probability of fixation. Thus, for neutrally evolving sites, the equation becomes the following: \( K = N_u \times \frac{1}{N} = u \). Nice . . .
OK; don’t sweat the details.

* The main points —
  * Hardy-Weinberg gives us a null model.
  * Mutation provides variation.
  * Mate choice REALLY matters.
  * Migration can make big differences.
  * Genetic drift can be counterintuitive, yet can be used to calculate lots of great stuff.
  * The neutral theory explains most variation in genomes, and gives us another null model.
Now that’s out of the way — modes of natural selection:

- **Directional selection** – one extreme phenotype is fittest, others are selected against. This is a type of positive selection.

- **Disruptive selection** – two or more extreme phenotypes are fitter than the intermediate phenotype, a.k.a. diversifying selection. This is also a type of positive selection.

- **Stabilizing selection** – extreme phenotypes are less fit than the optimal intermediate phenotype – most common in stable, unchanging environments, a.k.a. purifying or negative selection. Results in conservation of features.
These can be illustrated thus . . .
And a biggy — balancing selection.

- This is also known as balanced polymorphism and is often maintained through heterozygote advantage, sometimes called overdominance.

- It can also occur through frequency-dependent selection where the fitness of one phenotype depends on frequency of other phenotypes in the population.
The best known case is sickle cell anemia.

- Multiple alleles of a gene persist indefinitely in the population in balanced polymorphism.
- Why do seemingly harmful alleles persist?
- Often it’s heterozygote advantage:
  - The heterozygote has greater fitness than either homozygote.
- Sickle cell disease and malaria — those with both sickle cell alleles die, but those with only one, i.e. are heterozygotes, are resistant to malaria!
The distributions are congruent:
Several human diseases are maintained in the population because of heterozygote advantage:

<table>
<thead>
<tr>
<th>Person Who Has or Carries</th>
<th>Is Protected From</th>
<th>Possibly Because</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic fibrosis</td>
<td>Diarrheal disease</td>
<td>Carriers have too few functional chloride channels in intestinal cells, blocking toxin</td>
</tr>
<tr>
<td>G6PD deficiency</td>
<td>Malaria</td>
<td>Red blood cells inhospitable to malaria parasite</td>
</tr>
<tr>
<td>Phenylketonuria (PKU)</td>
<td>Miscarriage induced by ochratoxin A, a fungal toxin</td>
<td>Excess amino acid (phenylalanine) in carriers inactivates toxin</td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td>Malaria</td>
<td>Red blood cells inhospitable to malaria parasite</td>
</tr>
<tr>
<td>Tay-Sachs disease</td>
<td>Tuberculosis</td>
<td>Unknown</td>
</tr>
<tr>
<td>Noninsulin-dependent diabetes mellitus</td>
<td>Starvation</td>
<td>Tendency to gain weight protects against starvation during famine</td>
</tr>
</tbody>
</table>
That's enough on natural selection and population genetics for now.

Don't sweat the details! Just try to understand the big picture.

Next time macroevolution and phylogenetics.