Genetic variation: the raw material of evolution

Color pattern polymorphism in *Cepea* snails

• How much variation is there?

• How does novel variation arise?
Sources of phenotypic variation

1. Differences in genotype – Different genotypes produce different phenotypes

2. Differences in environment – Different environments produce different phenotypes

3. Interactions between genotype and environment – The relative values of phenotypes produced by different genotypes depend on the environment
Genetic variation

Environment 1

AA
Aa
aa
Environmental variation

Environment 1

AA  Aa  aa

Environment 2

AA  Aa  aa
Genotype × Environment variation

Environment 1

AA
Aa
aa

Environment 2

AA
Aa
aa

*** Although prevalent in nature, we will ignore the complication of G×E ***
It is GENETIC variation that is essential for evolution

- Selection can act on purely phenotypic variation

- But without genetic variation evolution will not occur
How much genetic variation is there?

1. Statistical analysis of quantitative traits

2. Studies at the molecular level
How much genetic variation is there?
Part I. Statistical analysis of quantitative traits

How much of this phenotypic variation is genetic?
Some basic statistics I: The mean

\[ \bar{x} = \sum_{i=1}^{n} f_i X_i \]

\[ \bar{x} = .25(4) + .5(8) + .25(12) = 8 \]

Where \( n \) is the number of different phenotype classes
Basic statistics II: The variance

\[ V = \sum_{i=1}^{n} f_i (X_i - \bar{x})^2 \]

Where \( n \) is the number of different phenotype classes

\[ V = .25(4-8)^2 + .5(8-8)^2 + .25(12-8)^2 = 8 \]
Basic statistics II: The variance

Population with variance = 8

\[ V = 0(0 - 8)^2 + 0.25(4 - 8)^2 + 0.5(8 - 8)^2 + 0.25(12 - 8)^2 + 0(16 - 8)^2 = 8 \]

Population with variance = 16

\[ V = 0.0625(0 - 8)^2 + 0.25(4 - 8)^2 + 0.375(8 - 8)^2 + 0.25(12 - 8)^2 + 0.0625(16 - 8)^2 = 16 \]
Using basic statistics to decompose phenotypic variation

\[ V_P = V_G + V_E \]

**Genetic Variance**

- Locations: 0, 4, 8, 12, 16
- Frequencies: 0.25, 0.5, 0.75, 1

**Environmental Variance**

- Locations: 0, -1, 0, 1, 0
- Variants: E1, E2, E3

**Phenotypic Variance**

- Locations: 0, 4, 8, 12, 16
- Frequencies: 0.25, 0.5, 0.75, 1
Genetic variation can be further decomposed

\[ V_G = V_A + V_I + V_D \]
What mechanisms contribute to each component?

Additive genetic variance ($V_A$) – Due to the additive effects of alleles

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>2</td>
</tr>
<tr>
<td>Aa</td>
<td>1</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
</tr>
</tbody>
</table>

Dominance variance ($V_D$) – Due to dominance

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>2</td>
</tr>
<tr>
<td>Aa</td>
<td>1</td>
</tr>
<tr>
<td>aa</td>
<td>2</td>
</tr>
</tbody>
</table>

Interaction variance ($V_I$) – Due to epistasis

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA (BB)</td>
<td>2</td>
</tr>
<tr>
<td>AA (Bb)</td>
<td>1</td>
</tr>
<tr>
<td>AA (bb)</td>
<td>2</td>
</tr>
</tbody>
</table>
It is **additive genetic variance** that determines the resemblance of parents and offspring.

How do we know how much additive genetic variation exists within a population?

**Additivity**

♂️  ♀️

Offspring need not look like parents!

**Epistasis or Dominance**

♂️  ♀️

Offspring need not look like parents!
The proportion of phenotypic variation that is genetic can be estimated by calculating “heritability”

- **Broad sense heritability** – Measures the proportion of phenotypic variation that is genetic

\[ H_B^2 = \frac{V_G}{V_G + V_E} = \frac{V_G}{V_P} \]

- **Narrow sense heritability** – Measures the proportion of phenotypic variation attributable to the additive action of genes. This is the measure relevant to N.S.

\[ h_N^2 = \frac{V_A}{V_A + V_I + V_D + V_E} \]

How can we measure narrow sense heritability?
One possibility is a parent-offspring regression

- The slope of the linear regression is an estimate of heritability

$$h^2 = \frac{\text{Cov}[z_{\text{Parent}}, z_{\text{Offspring}}]}{V[z_{\text{Parent}}]}$$

$$\text{Cov}[z_{\text{Parent}}, z_{\text{Offspring}}] = \frac{\sum_{i=1}^{n} (z_{\text{Parent},i} - \bar{z}_{\text{Parent}})(z_{\text{Offspring},i} - \bar{z}_{\text{Offspring}})}{n}$$
One possibility is a parent-offspring regression

- **Perfectly heritable** – Slope is 1.0

- **High heritability** – Slope is 0.8685

- **Low heritability** – Slope is 0.0756
### How heritable are most traits?

<table>
<thead>
<tr>
<th>Trait</th>
<th>Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk Yield in Cattle</td>
<td>0.3</td>
</tr>
<tr>
<td>Body length in pigs</td>
<td>0.5</td>
</tr>
<tr>
<td>Litter size in pigs</td>
<td>0.15</td>
</tr>
<tr>
<td>Wool length in sheep</td>
<td>0.55</td>
</tr>
<tr>
<td>Egg weight in chickens</td>
<td>0.6</td>
</tr>
<tr>
<td>Age at first laying</td>
<td></td>
</tr>
<tr>
<td>In chickens</td>
<td>0.5</td>
</tr>
<tr>
<td>Tail length in mice</td>
<td>0.6</td>
</tr>
<tr>
<td>Litter size in mice</td>
<td>0.15</td>
</tr>
</tbody>
</table>

After Falconer (1981)

For almost any trait ever measured, there is abundant additive genetic variation!
A limitation of the statistical approach

Can never accurately reveal how many genetic loci are responsible for observed levels of variation
How much genetic variation is there?
Part II: Molecular variability

• Prior to 1966, it was generally assumed that populations were, in large part, genetically uniform.

• In 1966, two landmark papers (Lewinton and Hubby, 1966; Harris, 1966) turned this conventional wisdom on its head, demonstrating an abundance of genetic polymorphism.
So what did these landmark studies really show?

**Genetic polymorphism** – The presence of two or more alleles in a population, with the rarer allele having a frequency greater than .01.

Using protein gel electrophoresis, these studies showed that roughly 1/3 of all loci are polymorphic in both humans and *Drosophila*. 
Subsequent studies found the same thing!

<table>
<thead>
<tr>
<th></th>
<th>NUMBER OF SPECIES EXAMINED</th>
<th>AVERAGE NUMBER OF LOCI PER SPECIES</th>
<th>AVERAGE PROPORTION OF LOCI POLYMORPHIC PER POPULATION</th>
<th>HETEROZYGOUS PER INDIVIDUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drosophila</td>
<td>28</td>
<td>24</td>
<td>0.529</td>
<td>0.150</td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
<td>18</td>
<td>0.531</td>
<td>0.151</td>
</tr>
<tr>
<td>Haplodiploid wasps</td>
<td>6</td>
<td>15</td>
<td>0.243</td>
<td>0.062</td>
</tr>
<tr>
<td>Marine invertebrates</td>
<td>9</td>
<td>26</td>
<td>0.587</td>
<td>0.147</td>
</tr>
<tr>
<td>Marine snails</td>
<td>5</td>
<td>17</td>
<td>0.175</td>
<td>0.083</td>
</tr>
<tr>
<td>Land snails</td>
<td>5</td>
<td>18</td>
<td>0.437</td>
<td>0.150</td>
</tr>
<tr>
<td>Fishes</td>
<td>14</td>
<td>21</td>
<td>0.306</td>
<td>0.078</td>
</tr>
<tr>
<td>Amphibians</td>
<td>11</td>
<td>22</td>
<td>0.336</td>
<td>0.082</td>
</tr>
<tr>
<td>Reptiles</td>
<td>9</td>
<td>21</td>
<td>0.231</td>
<td>0.047</td>
</tr>
<tr>
<td>Birds</td>
<td>4</td>
<td>19</td>
<td>0.145</td>
<td>0.042</td>
</tr>
<tr>
<td>Rodents</td>
<td>26</td>
<td>26</td>
<td>0.202</td>
<td>0.054</td>
</tr>
<tr>
<td>Large mammals</td>
<td>4</td>
<td>40</td>
<td>0.233</td>
<td>0.037</td>
</tr>
<tr>
<td>Plants</td>
<td>8</td>
<td>8</td>
<td>0.464</td>
<td>0.170</td>
</tr>
</tbody>
</table>

Source: Futuyma, *Evolutionary Biology, 3’rd Edition*
Suggests that almost every individual in a sexually reproducing species is genetically unique!

- Even with only two alleles per locus, the estimated 3000 polymorphic loci in humans could generate $3^{3000} = 10^{1431}$ different genotypes!
The bottom line:

No matter how you cut it, there is abundant genetic variation WITHIN populations, and thus ample opportunity for selection to act.
Assessing genetic variation and Hardy-Weinberg I: a practice problem

The scenario: A group of biologists was studying a population of elk in an effort to quantify genetic variation at disease resistance locus. Through DNA sequencing, the biologists have determined that there are two alleles at this locus, A and a. Sequencing analysis of many individuals has also allowed the frequency of the alleles and the corresponding diploid genotypes to be estimated.

The data:
Frequency of the A allele is $p = 0.4$
Frequency of the a allele is $q =$ ?

Frequency of the AA genotype is: 0.06
Frequency of the Aa genotype is: 0.80
Frequency of the aa genotype is: 0.14

The question:
Is this population in Hardy-Weinberg Equilibrium? Justify your response.
Increasing the scale:
Genetic variation among populations

Genetic variation within a single population

```
Aa  Aa  Aa
AA  AA  AA
Aa  AA  Aa
AA  Aa  AA
```

Genetic variation among populations

```
AA  AA  AA
AA  AA  AA
AA  AA  AA
```

```
aa  aa  aa
aa  aa  aa
aa  aa  aa
```
Genetic variation among populations

Genetic variation in human resistance to Malaria
Increasing the scale: Genetic variation among species

Chum < 32 lbs  
Chinook < 100 lbs  
Coho < 26 lbs  
Pink < 12 lbs  
Sockeye < 16 lbs

These different species are genetically differentiated with respect to adult size.
We now know that genetic variation is hierarchical
An applied problem: genetic variation and conservation

Sockeye Salmon

Redfish Lake, Idaho

Populations vs. Species: Which is more relevant?
Assessing genetic variation and Hardy-Weinberg II: a practice problem

The scenario: A group of biologists is studying a population of flowers where flower color is controlled by a single diploid locus with two alleles. Individuals with genotype AA make white flowers, individuals with genotype Aa make red flowers, and individuals with genotype aa make red flowers.

The data:
Frequency of the white flowers is \( f(white) = 0.4 \)
Frequency of red flowers is \( f(red) = ? \)

The questions:
1. Which allele, \( A \) or \( a \) is dominant?
2. Assuming that this population is in Hardy-Weinberg Equilibrium, what is the frequency of the \( A \) allele?
3. Assuming that this population is in Hardy-Weinberg Equilibrium, what is the frequency of the \( a \) allele?
Where does genetic variation come from?

1. Mutation – An alteration of a DNA sequence that is inherited

2. Recombination – The formation of gametes with combinations of alleles different from those that united to form the individual that produced them.

3. Gene flow – The incorporation of genes into the gene pool of one population from one or more other populations.

4. Hybridization – The incorporation of genes into the gene pool of one species from another species.
Important facts about mutation

• Mutations are RANDOM with respect to fitness

• Only mutations that are inherited (germline) are relevant to evolution
Estimating the mutation rate

• Direct methods – Simply counting new mutations

• Statistical methods – Based on increases in phenotypic variance
Direct estimation of the mutation rate

\[ G = 1 \]

50,000 flies all homozygous for the (hypothetical) recessive red eye allele (A)

\[ G = 2 \]

50,000 flies but now 1 has white eyes indicating genotype (Aa)

We could then estimate that the per locus mutation rate as \(1/100,000 = .00001\)
Implications of these estimates for mutation rates

• As a gross average, the per locus mutation rate is $10^{-6} - 10^{-5}$ mutations per gamete per generation.

• As a gross average, humans have 150,000 functional genes

• $10^{-5} \times 150,000 = 1.5$

This suggests that EVERY gamete carries a new, phenotypically detectable mutation somewhere in its genome!!!
Spontaneous mutation rates of specific genes detected by phenotypic effects

<table>
<thead>
<tr>
<th>Species</th>
<th>Taxonomic group</th>
<th>Number of mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. Coli</em></td>
<td>Bacteria</td>
<td>0.0025</td>
</tr>
<tr>
<td><em>S. acidocaldarius</em></td>
<td>Archaea</td>
<td>0.0018</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>Fungi</td>
<td>0.0030</td>
</tr>
<tr>
<td><em>S. cerevisiae</em></td>
<td>Fungi</td>
<td>0.0027</td>
</tr>
<tr>
<td><em>C. elegans</em></td>
<td>Roundworms</td>
<td>0.0360</td>
</tr>
<tr>
<td><em>D. Melanogaster</em></td>
<td>Insects</td>
<td>0.1400</td>
</tr>
<tr>
<td><em>M. Musculus</em></td>
<td>Mammals</td>
<td>0.9000</td>
</tr>
<tr>
<td><em>H. sapiens</em></td>
<td>Mammals</td>
<td>1.6000</td>
</tr>
</tbody>
</table>

Statistical estimation of new mutational genetic variance

Inbreed until all additive genetic variance for some trait of interest is lost

Mate at random and measure heritability

It would then take only 100 generations for $h^2$ to equal .1!
What effect do mutations have on fitness?

Although we know very little, we know that new mutations are generally deleterious.
Recombination as a source of variation

Recombination absent

Zygotes
AB/AB
ab/AB
ab/ab
AB/ab

Meiosis (no recombination)

Gametes
AB
ab

Fusion

Zygotes
AB/AB
ab/AB
ab/ab
AB/ab

Recombination present

Zygotes
AB/AB
ab/AB
ab/ab
AB/ab

Meiosis (recombination)

Gametes
Ab
AB
ab
aB

Fusion

Zygotes
Ab/AB
aB/AB
ab/AB
AB/ab

Recombination generates new COMBINATIONS of genes
Gene flow as a source of variation

Population 1

\[ V_G = 0 \]

AA  AA
AA  AA
AA  AA
AA

Population 2

\[ V_G = 0 \]

aa  aa
aa  aa
aa

Population 1

\[ V_G > 0 \]

AA  Aa
AA  aA
AA  aa
aa

Population 2

\[ V_G > 0 \]

AA  aa
aA  aa
Aa
Hybridization as a source of genetic variation

• Hybridization reshuffles genes between species

• Often has dramatic phenotypic effects

• *IF* offspring are viable and fertile, hybridization can be an important source of new genetic variation
Hybridization as a source of genetic variation

Aquilegia formosa
Lower elevations (6000-10,000 ft)

Aquilegia pubescens
High elevations (10,000-13,000 ft).

Both species grow in the Sierra Nevada mountains of California
Hybridization as a source of genetic variation

Formosa - Pubescens hybrid zone
Summary

• There is abundant genetic variation in natural populations

• Mutation is the ultimate source of genetic variation

• Recombination, gene flow, and hybridization redistribute genetic variation
Practice Problem

You are studying a population of Steelhead Trout and would like to know to what extent body mass is heritable. To this end, you measured the body mass of male and female Steelhead as well as the body mass of their offspring. Use the data from this experiment (below) to estimate the heritability of body mass in this population of Steelhead.

<table>
<thead>
<tr>
<th>Maternal Body Mass (Kg)</th>
<th>Paternal Body Mass (Kg)</th>
<th>Average Offspring Body Mass (Kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>2.5</td>
<td>2.9</td>
<td>2.5</td>
</tr>
<tr>
<td>1.9</td>
<td>3.1</td>
<td>2.7</td>
</tr>
<tr>
<td>2.2</td>
<td>2.8</td>
<td>2.4</td>
</tr>
<tr>
<td>1.8</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td>2.4</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>2.3</td>
<td>2.9</td>
<td>2.7</td>
</tr>
</tbody>
</table>