

Genetic and Phenotypic Variation

I) Introduction

The goal of population genetics is to understand and explain the origin and maintenance of genetic variation within natural populations.

A) Types of Genetic Variation

- 1) **SPORTS** – sudden, rare, spontaneous mutations (albinos)
- 2) **Conspicuous Polymorphisms** – commonly observed differences such as color in Biston (pepper moths), Lady bugs, etc.
- 3) **Chromosomal Variation** – usually no morphological differences but inverted, fused or otherwise rearranged. Can cause fitness reduction in heterozygotes. (Dobzhansky 1941)
- 4) **Biochemical** – differences in allozyme (protein) mobility (Lewontin and Hubby 1966) or blood groups.
- 5) **DNA variation** – alterations in the sequence of DNA
 - a) **RFLP**
 - b) **Sequencing**
- 6) **Continuous variation (i.e., quantitative)** – pretty much everything else. Has genetic and environmental components

B) Views of naturally occurring variation (Dobzhansky 1955)

- 1) **Classical School** – populations are largely monomorphic (i.e., homozygous) occasionally new, rare variants arise

A	B	C	D	E	F	G	H
<hr/>							
A	B	C	d	E	F	G	H

Wild type and mutant heterozygote. If the new mutant is more fit than the wild type it will go to fixation (i.e., 100% frequency in the population) otherwise it will be purged.

- 2) **Balanced school** – populations are highly polymorphic and most individuals are heterozygous. Variants are maintained by balancing selection (i.e., heterozygotes are better fit than homozygotes).

A	b	c''	d'	E	F	g''	h'
<hr/>							
a	B	C	d	e'	f	g'	h'''

C) Influences on Genetic Variation

- 1) **Mutation** – generally increases variation (even if only briefly; assuming an infinite allele model)
- 2) **Drift** – random changes in allele frequencies over time due to “sampling effect.” Generally reduces variation since rare alleles can be easily lost

- 3) **Selection – differential survival and reproduction of a genotype. Can increase, decrease or stabilize the frequency of variation.**
- 4) **Migration – same as 3**
- 5) **Breeding System – same as 3**
- 6) **Mostly all of the above occur at the same time**

D) Ways To Reveal Genetic Variation – nearly unlimited

- 1) **Protein Polymorphisms (OH 2.1)**
- 2) **Restriction site differences (OH 2.2 & OH 2.3)**
- 3) **DNA sequencing (OH 2.4)**

E) Ways to measure Genetic Variation

- 1) **Frequency – most basic. Can be used for almost any discrete marker (i.e., alleles) (OH 2.5)**
- 2) **% Polymorphic loci – will depend on sample size (OH 2.6)**

$$= \frac{\text{\#loci polymorphic}}{\text{\# loci scored}}$$

- a. **Include all loci – commonly done for allozyme studies**
 - b. **Include only polymorphic loci – common for DNA studies**
- 3) **Heterozygosity – can be informative in understanding breeding patterns, evolutionary history, population subdivision, migration, etc. can be observed or expected**

$$\text{direct count} = \frac{\text{\# heterozygotes}}{\text{Total \# individuals}}$$

$$E(H) = 1 - \sum_{i=1}^m p_i^2$$

m = number of alleles (OH 2.7)

- 4) **Number of alleles per locus – less informative for allozyme loci but more so for DNA. Information not immediately available from % polymorphic or heterozygosity.**
- 5) **Nucleotide diversity (π) – the average proportion of nucleotide differences between all possible pairs of sequences in the sample**

\hat{X}_i = population frequency of i^{th} sequence (haplotype)

\hat{X}_j = population frequency of j^{th} sequence (haplotype)

N = total number of sequences (not just unique)

π_{ij} = proportion of different nucleotides between the sequences.

P_{ij} = number of differences between sequence i and j .

m_T = length of sequence

n = number of unique sequences

$$\pi_{ij} = \frac{P_{ij}}{m_T}$$

OH 2.8

$$\hat{\pi} = \frac{N}{N-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{X}_i \hat{X}_j \pi_{ij}$$

$$\text{var}(\hat{\pi}) = \frac{N+1}{2(N-1)m_T} \hat{\pi} + \frac{2(N^2 + N + 3)}{9N(N-1)} \hat{\pi}^2$$

- 6) Nucleotide polymorphism (θ) – proportion of sites expected to be variable (i.e., heterozygosity). Effected by sample size both in # of individuals and length of sequence.

P_n = # polymorphic Sites

m_T = total number of sites

n = number of unique sequences

$$S_n = \frac{P_n}{m_T}$$

$$a_i = \frac{n-1}{i}$$

$$\hat{\theta} = \frac{S_n}{a_i} \quad \text{OH 2.9}$$

$$\text{Var}(\hat{\theta}) = \frac{\hat{\theta}}{m_T a_i} + \frac{a_2 \hat{\theta}^2}{a_i^2}$$

$$\text{where } a_2 = \frac{n-1}{i^2}$$

F) How much variation is there in natural populations?

- 1) A lot and a little (OH 2.10)