

POPULATION GENETICS:

- The study of the rules governing the maintenance and transmission of genetic variation in natural populations.

DARWINIAN EVOLUTION BY NATURAL SELECTION

- Many more individuals are born than survive (**COMPETITION**).
- Individuals within species are variable (**VARIATION**).
- Some of these variations are passed on to offspring (**HERITABILITY**).
- Survival and reproduction are **not** random. There must be a correlation between fitness and phenotype.

The "rediscovery" of Mendel's genetic studies in 1902 by William Bateson completed the missing model for the inheritance of genetic factors.

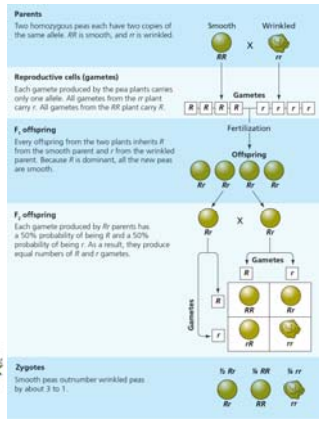
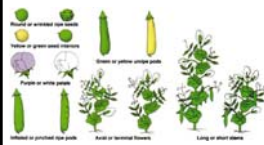


Gregor Mendel

- Mendel published his work in the Transactions of the Brunn Society of Natural History in 1866.

Mendel's Laws:

- Segregation
- Independent assortment



SEXUAL REPRODUCTION CONTRIBUTES TO VARIATION

Example – A Line Cross Experiment

- Consider 2 diploid individuals with 3 loci and 2 alleles,

Parents: aabbcc x AABbCc

F₁ progeny: AaBbCc

F₂ progeny:

AABBCC	AABBcc	AABbCc	27 COMBINATIONS
AABbCC	AABbcc	AABbCc	
AAbbCC	AAbbCc	AAbbCc	
AaBBCC	AaBBcc	AaBBCc	
AaBbCC	AaBbcc	AaBbCc	
AabbCC	Aabbcc	AabbCc	
aaBBCC	aaBBcc	aaBBCc	
aaBbCC	aaBbcc	aaBbCc	
aabbCC	aabbcc	aabbCc	

Mechanisms of Evolution: Mendelian Genetics in Populations

- Genetic variation is the raw material of evolutionary change: how do we measure it?
- What are the forces that cause genetic changes within populations? That is, what mechanisms cause evolutionary change?

Population Genetics

- Evolution can be defined as a change in gene frequencies through time.
- Population genetics tracks the fate, across generations, of Mendelian genes in populations.
- Population genetics is concerned with whether a particular allele or genotype will become more or less common over time, and **WHY**.

A few things to keep in mind as we take an excursion into population genetic theory:

"Make things as simple as possible, but no simpler."
---Einstein

"All models are wrong, some are useful."
---Box

"No theory should fit all the facts because some of the facts are wrong."
---Bohr

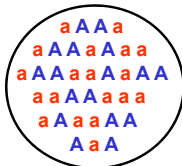
"Before I came here I was confused about this subject. Having listened to your lecture I am still confused. But on a higher level."
---Fermi

Assumptions:

- Diploid, autosomal locus with 2 alleles: **A** and **a**
- Simple life cycle:

PARENTS (DIPLOID) → GAMETES (HAPLOID) → ZYGOTES (DIPLOID)

These parents produce a large gamete pool (Gene Pool) containing alleles **A** and **a**.



Gamete (allele) Frequencies:

$$\text{Freq}(A) = p$$
$$\text{Freq}(a) = q$$

$$\Rightarrow p + q = 1$$

Genotype Frequencies of 3 Possible Zygotes:

AA Aa aa

$$\text{Freq}(AA) = p_A \times p_A = p_A^2$$

$$\text{Freq}(Aa) = (p_A \times q_a) + (q_a \times p_A) = 2p_A q_a$$

$$\text{Freq}(aa) = q_a \times q_a = q_a^2$$

$$\Rightarrow p^2 + 2pq + q^2 = 1$$

General Rule for Estimating Allele Frequencies from Genotype Frequencies:

Genotypes: AA Aa aa

Frequency: p^2 $2pq$ q^2

\Rightarrow Frequency of the **A** allele:

$$p = p^2 + \frac{1}{2}(2pq)$$

\Rightarrow Frequency of the **a** allele:

$$q = q^2 + \frac{1}{2}(2pq)$$

Sample Calculation: Allele Frequencies

Assume N = 200 indiv. in each of two populations 1 & 2

- Pop 1 : 90 AA 40 Aa 70 aa
- Pop 2 : 45 AA 130 Aa 25 aa

In Pop 1 :

- $p = p^2 + \frac{1}{2}(2pq) = 90/200 + \frac{1}{2}(40/200) = 0.45 + 0.10 = \mathbf{0.55}$
- $q = q^2 + \frac{1}{2}(2pq) = 70/200 + \frac{1}{2}(40/200) = 0.35 + 0.10 = \mathbf{0.45}$

In Pop 2 :

- $p = p^2 + \frac{1}{2}(2pq) = 45/200 + \frac{1}{2}(130/200) = 0.225 + 0.325 = \mathbf{0.55}$
- $q = q^2 + \frac{1}{2}(2pq) = 25/200 + \frac{1}{2}(130/200) = 0.125 + 0.325 = \mathbf{0.45}$

Main Points:

- $p + q = 1$ (more generally, the sum of the allele frequencies equals one)
- $p^2 + 2pq + q^2 = 1$ (more generally, the sum of the genotype frequencies equals one)
- Two populations with markedly different genotype frequencies can have the same allele frequencies

The Hardy-Castle-Weinberg Law

- A single generation of random mating establishes H-W equilibrium genotype frequencies, and neither these frequencies nor the gene frequencies will change in subsequent generations.

$$p^2 + 2pq + q^2 = 1$$



Hardy

H-W ASSUMPTIONS:

- 1) Mating is random (with respect to the locus).
- 2) The population is infinitely large. (no sampling error – Random Genetic Drift)
- 3) Genes are not added from outside the population (no gene flow or migration).
- 4) Genes do not change from one allelic state to another (no mutation).
- 5) All individuals have equal probabilities of survival and reproduction (no selection).

IMPLICATIONS OF THE H-W PRINCIPLE:

- 1) A random mating population with no external forces acting on it will reach the equilibrium H-W frequencies in a single generation, and these frequencies remain constant there after.
- 2) Any perturbation of the gene frequencies leads to a new equilibrium after random mating.
- 3) The amount of **heterozygosity** is maximized when the gene frequencies are intermediate.

$2pq$ has a maximum value of 0.5 when $p = q = 0.5$

GENOTYPE VERSUS GENE FREQUENCIES

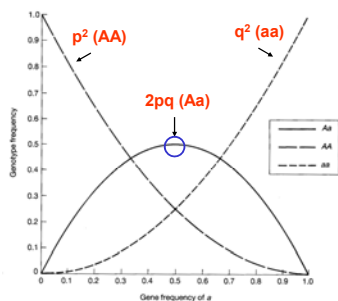


Figure 5.2 Hardy-Weinberg proportions of genotypes AA, Aa, and aa in relation to the frequency of the gene a (q)

FOUR PRIMARY USES OF THE H-W PRINCIPLE:

- 1) Enables us to compute genotype frequencies from generation to generation, even with selection.
- 2) Serves as a **null** model in tests for natural selection, nonrandom mating, etc., by comparing observed to expected genotype frequencies.
- 3) Forensic analysis.
- 4) Expected heterozygosity provides a useful means of summarizing the molecular genetic diversity in natural populations.

Genotype frequencies for Wing-color Polymorphism in a Natural Population of the Moth *Panaxia dominula* (Fisher & Ford, 1947)

Color Pattern	<i>dominula</i>	<i>medionigra</i>	<i>bimaculata</i>	Total
Genotype	B ₁ B ₁	B ₁ B ₂	B ₂ B ₂	
Sample Size (N _i)	905	78	3	N = 986
Frequency (P _i)	0.918	0.079	0.003	1.000

Allele Frequencies:

$$p_1 = p_{11} + \frac{1}{2} p_{12} = 0.918 + \frac{1}{2} (0.079) = 0.958$$

$$p_2 = 1 - p_1 = 0.041$$

Expectations (\hat{N}_i)	$p_1^2 N$	$2p_1p_2 N$	$p_2^2 N$	
	905	79	2	

EVOLUTIONARY THOUGHT AFTER DARWIN

- By the 1870s, most scientists accepted the historical reality of evolution (and this has been true ever since).
- It would be at least 60 years after the publication of *The Origin of Species* before natural selection would come to be widely accepted.
- People wanted life itself to be purposeful and creative, and consequently did not find natural selection appealing.
 - Neo-Lamarckism -- inheritance of acquired characteristics.
 - Orthogenesis -- variation that arises is directed toward a goal.
 - Mutationism -- discrete variations are all that matter.

MUTATIONISM AND THE IMPACT OF MENDEL

- Gregor Mendel's research was published in 1866, but was not noticed until 1900.
- NOTE: Darwin knew nothing about the mechanism of inheritance when he conceived of natural selection.
- MUTATIONIST THEORIES (based on Mendel's work):
 - Emphasized the importance of DISCRETE VARIATION
 - T. H. Morgan -- the founder of *Drosophila* genetics.
 - R. Goldschmidt -- theory of "hopeful monsters".

Mendelian genetics disproved Lamarckian and blending inheritance theories. However, the emphasis on discrete variation caused a rift between mutation and natural selection that was initially damaging to the field.

THE MODERN SYNTHESIS OF EVOLUTIONARY BIOLOGY

- The rift between genetics and natural selection was resolved in the 1930s and 1940s.
- This synthesis was forged from the contributions of geneticists, systematists, and paleontologists.

The Genetical Theory of Natural Selection 1930

- Fisher united Mendelian population genetics with the inheritance of continuous traits.



"The Correlation between Relatives on the Supposition of Mendelian Inheritance"
Philosophical Transactions of the Royal Society of Edinburgh in 1918, (volume 52, pages 399–433).

Sir R. A. Fisher
1890-1962



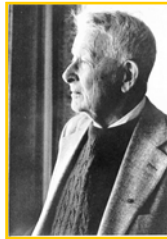
The Causes of Evolution 1932

- Developed the mathematical theory of gene frequency change under selection (and many other interesting applications).

J. B. S. Haldane
1892-1964



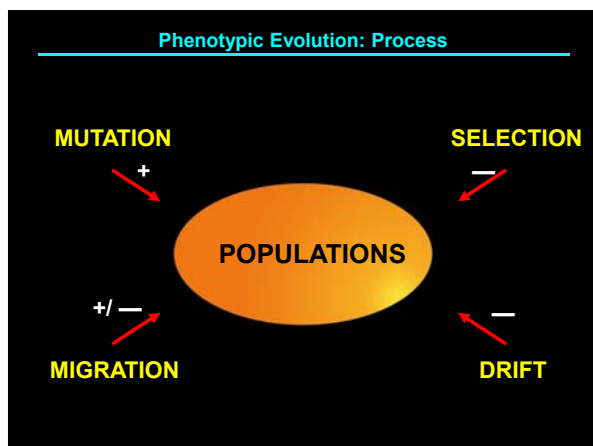
- Developed the mathematical framework for understanding the genetic consequences of migration, effective population size, population subdivision, and conceived of the concept of adaptive landscapes.



Sewall Wright
On his 90th birthday
1889-1988

Outcomes of the “MODERN SYNTHESIS”

- Populations contain genetic variation that arises by random mutation.
- Populations evolve by changes in gene frequency.
- Gene frequencies change through random genetic drift, gene flow, and natural selection.
- Most adaptive variants have small effects on the phenotype so changes are typically gradual.
- Diversification comes about through speciation.



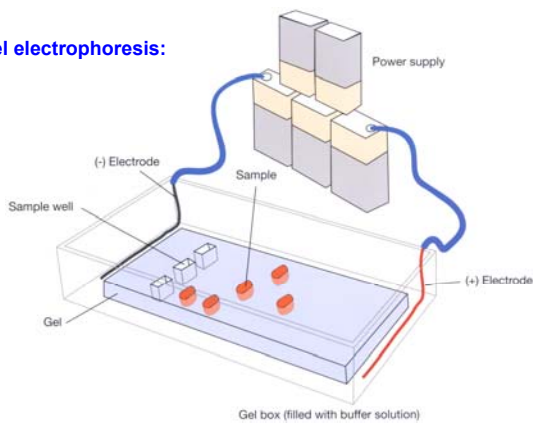
MEASURING GENETIC VARIATION IN NATURAL POPULATIONS

TWO COMMONLY USED MEASURES TO QUANTIFY GENETIC VARIATION ARE:

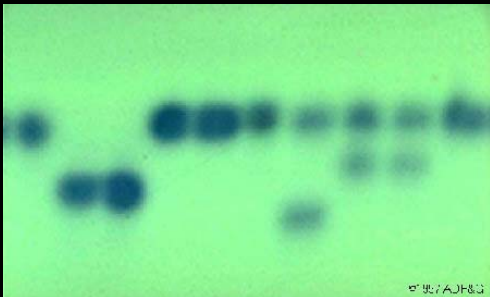
P – the proportion of polymorphic loci (those that have 2 or more alleles)

H – the average heterozygosity = proportion of loci at which a randomly chosen individual is heterozygous.

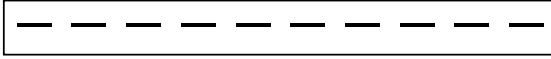
Gel electrophoresis:



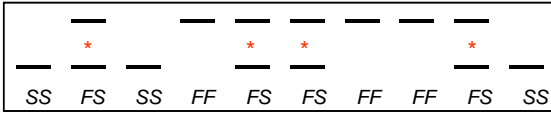
AN ALLOZYME GEL



Monomorphic locus: all individuals fixed for same allele



Polymorphic locus: 2 alleles



- Frequency of *F* allele, $p = 10 / 20 = 0.5$
- Frequency of *S* allele, $q = 10 / 20 = 0.5$
- $H = 4 / 10 = 0.4$

ALLOZYME VARIATION IN MAJOR TAXA

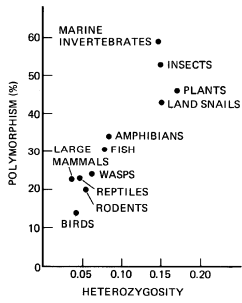
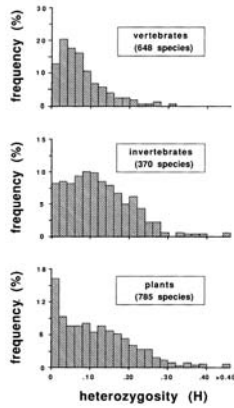


FIGURE 5. Genic variation in some major groups of animals and in plants. Mean values of P and H are plotted. (See Table V.)

In F. J. Ayala (ed.) 1976. Molecular Evolution. Sinauer Assoc.

ALLOZYME VARIATION WITHIN MAJOR TAXA



ALLOZYME VARIATION AMONG *Drosophila* spp.

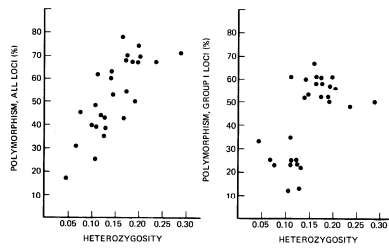
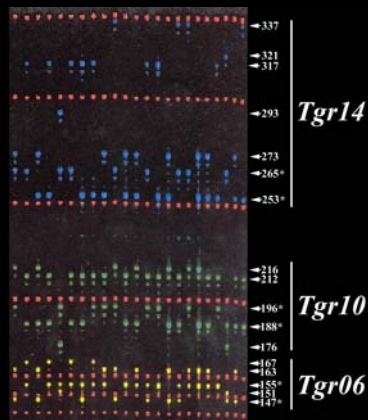


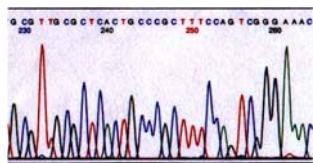
FIGURE 4. Mean heterozygosity (H) over all loci examined in relation to polymorphic proportion of all loci (left) and of group I loci (right) for 28 species of *Drosophila*.

In F. J. Ayala (ed.) 1976. Molecular Evolution. Sinauer Assoc.

3 MICROSATELLITE
LOCI, SHOWING
HIGH LEVELS OF
POLYMORPHISM
WITHIN A
POPULATION.



SEQUENCE DATA



- The definitive source of information regarding genetic variation.
- Usually obtained by PCR and direct sequencing of a particular gene of interest.
- Sequence-level variation can also be assessed using restriction enzymes (RFLPs).
- More on sequences when we get to "molecular evolution".

THE EXTENSION OF H-W TO LOCI WITH MANY ALLELES

- Regardless of the number of alleles per locus, the H-W principle still applies as long as the organisms are diploid.
- Assume that you have a locus with n alleles (A_1, A_2, \dots, A_n), and the allele frequencies are given by $p_1, p_2, p_3, \dots, p_n$.
- The expected frequency of any homozygote is just the square of the allele frequency.

$$F(A_1A_1) = p_1^2, F(A_6A_6) = p_6^2, \text{ etc.}$$

- The expected frequency of any heterozygote is 2 times the product of the respective allele frequencies.

$$F(A_1A_3) = 2p_1p_3, F(A_3A_5) = 2p_3p_5, \text{ etc.}$$

- Expected heterozygosity is $1 - \sum(p_i^2)$.

DNA POLYMORPHISM AS A TOOL FOR FORENSICS???

With the recognition... that the... genome is replete with DNA sequence polymorphisms such as RFLP's, it was only a small leap to imagine that DNA could, in principle, provide the ultimate identifier.

E. S. Lander, 1991

DNA analysis will be to the end of the 20th century what fingerprinting was to the 19th.

Melson, 1990

APPLICATIONS OF H-W TO FORENSICS

▪Example:

Defendant Genotype

Tissue Sample From
Crime Scene

$A_{14}A_{27}$

$A_{14}A_{27}$

▪Question:

- What is the probability of this match being a false positive???
- In other words what is the probability that this match occurred simply by chance?

PROBABILITY OF A FALSE POSITIVE

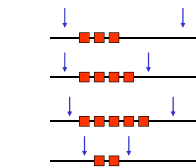
Frequency of A_{14} in population = p_{14}
 Frequency of A_{27} in population = p_{27}

- Expected fraction of random individuals in the population with the same genotype as the defendant = $2 p_{14} p_{27}$

PRODUCT RULE:

- With multiple loci, the probability of a false positive is simply equal to the product of the locus-specific probabilities.
- As this probability becomes small, we become more confident that the match is real.

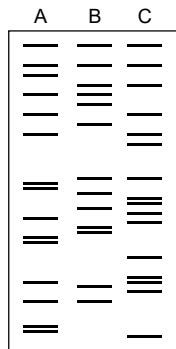
VNTR Loci:



Highly repetitive blocks of DNA (■) found through out the genome.

DNA is cut with restriction enzymes (↓), labeled with ^{32}P and run through a gel to reveal a characteristic banding pattern

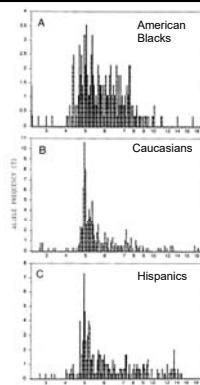
INDIVIDUAL



Genetic diversity of VNTR Loci in Human Populations

Average heterozygosity of 6 VNTR loci in ethnic groups:

Caucasian	0.89
Chinese	0.84
Cambodian	0.80
Melanesian	0.76
Australian aborigine	0.77



FROM: Balazs et al. 1989, Am. J. Hum. Genet. 44:182-190.
 Balazs et al. 1992, Genetics 131:191-198

Figure 4 Allele frequency distribution of 10 VNTR loci in three ethnic groups. The number of alleles analyzed for these figures were 12, 282 for American Blacks, 191 for Caucasians, and 173 for Hispanics.

Some Definitions:

- **Population:** A freely interbreeding group of individuals.
- **Gene Pool:** The sum total of genetic information present in a population at any given point in time.
- **Phenotype:** A morphological, physiological, biochemical, or behavioral characteristic of an individual organism.
- **Genotype:** The genetic constitution of an individual organism.
- **Locus:** A site on a chromosome, or the gene that occupies the site.
- **Gene:** A nucleic acid sequence that encodes a product with a distinct function in the organism.
- **Allele:** A particular form of a gene.
- **Gene (Allele) Frequency:** The relative proportion of a particular allele at a single locus in a population (a number between 0 and 1).
- **Genotype Frequency:** The relative proportion of a particular genotype in a population (a number between 0 and 1).
