Evolution and Medicine

- Evolution of the Genomes

- **Genetic Diversity and Polymorphism**

- Evolutionary Mechanisms
  - Microevolution vs Macroevolution
  - Population Genetics
  - Evolution and Medicine

Personalized genomics

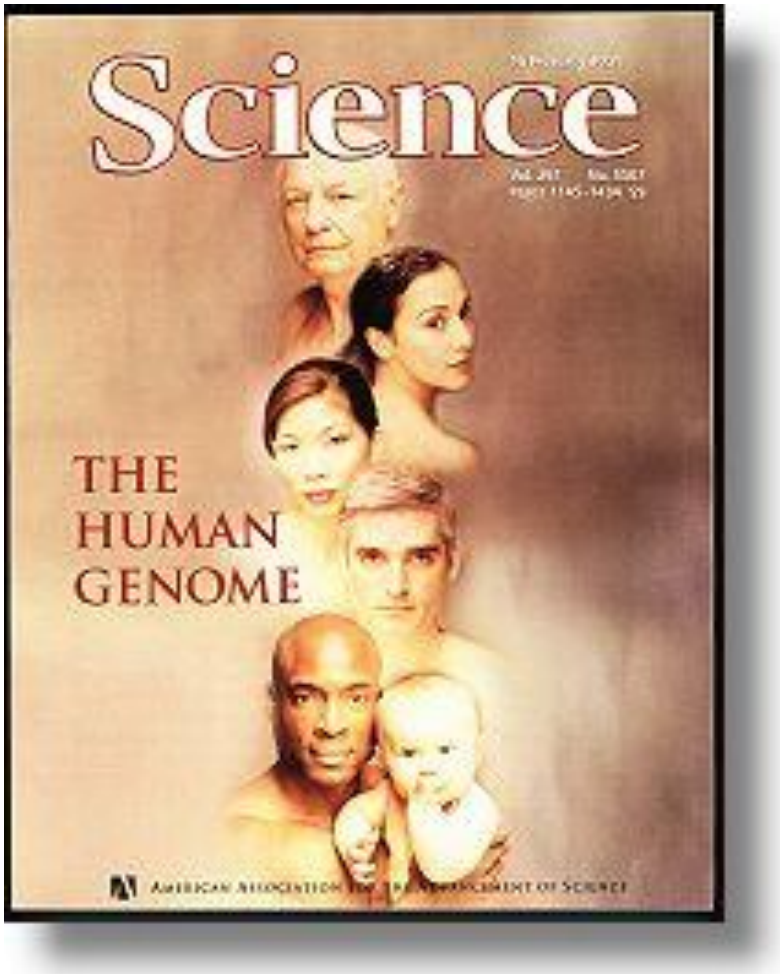
When in Doubt, Spit It Out

From an individual to a population

- It took a long time (20-25 yrs) to produce the draft sequence of the human genome.
- Soon (within 5-10 years), entire populations can have their DNA sequenced. Why do we care?
- Individual genomes vary by about 1 in 1000bp.
- These small variations account for significant phenotype differences.
  - Disease susceptibility.
  - Response to drugs are just few but important examples
- How can we understand genetic variation in a population, and its consequences?

Population Genetics

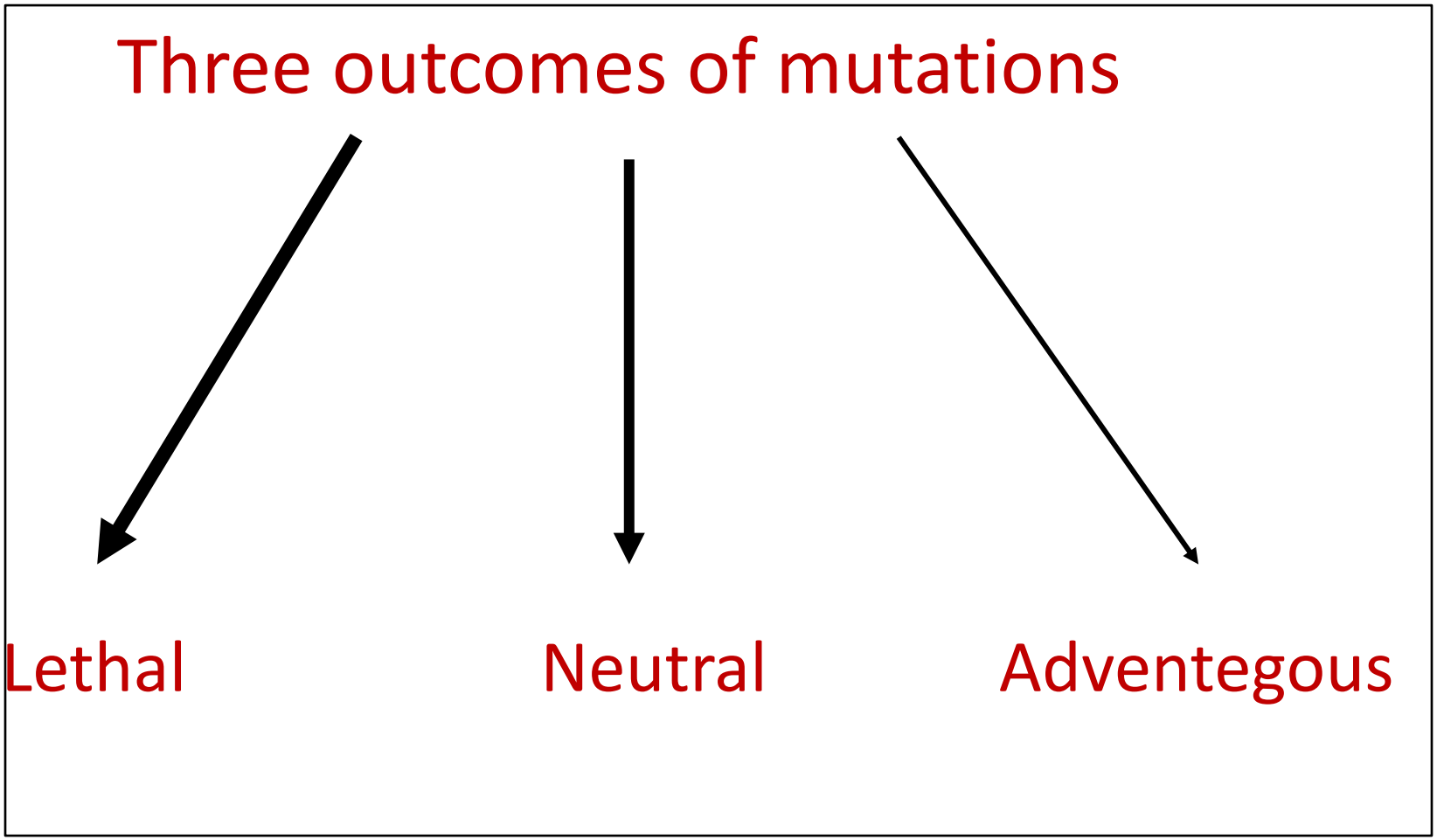
- Individuals in a species (population) are phenotypically different.
- These differences are inherited (genetic). Understanding the genetic basis of these differences is a key challenge of biology!
- The analysis of these differences involves many interesting algorithmic questions.
- We will use these questions to illustrate algorithmic principles, and use algorithms to interpret genetic data.



Variation in DNA

- The DNA is inherited by the child from its parent.
- The copying is not identical, but might be mutated.
- If the mutation lies in a gene,….

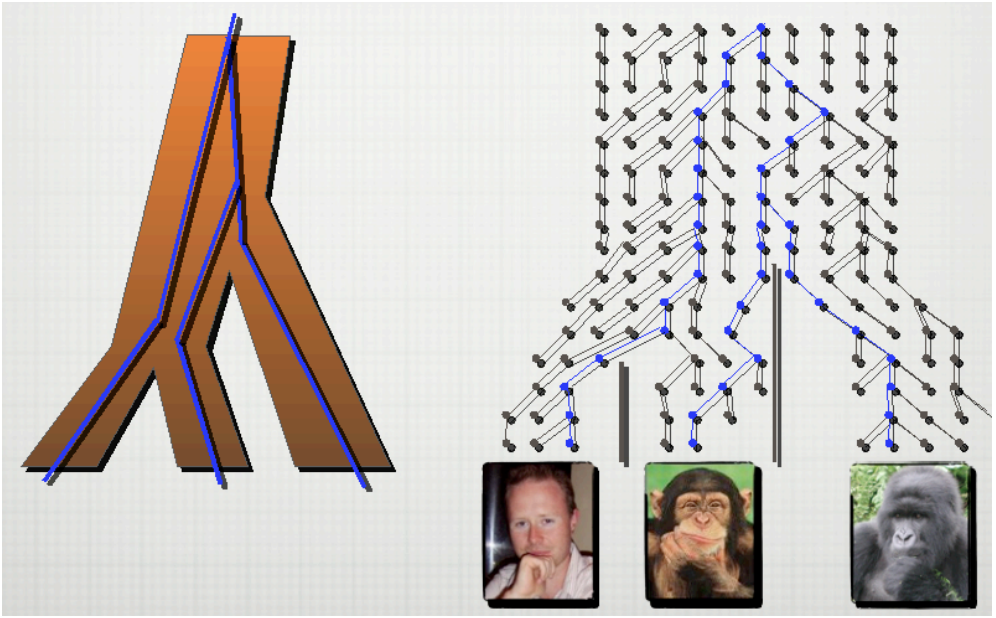
  - Different proteins are produced but **not only proteins**
  - Genes are switched on or off
  - Different phenotype.



- Most mutations are selectively neutral!
- They drift in the population, eventually getting eliminated, or fixed by random chance.

(Comparative) Genomics vs. (Population) Genetics

- Mutations accumulate over time.
- In looking at the DNA of different species
  - DNA has had a lot of time to mutate, and is not expected to be identical.
  - If the DNA is highly similar, that region is functionally important.
- In comparing DNA from a population
  - Not enough time to mutate, so DNA is expected to be identical.
  - The few differences that exist mediate phenotypic differences.



Population genetics

- By sampling DNA from a population, we can answer the following

  - What are the sources of variation?
  - As mutations arise, they are either neutral and subject to evolutionary drift, or they are (dis-)advantageous and under selective pressure. Can we tell?
  - If you had DNA from many sub-populations, Asian, European, African, can you separate them?
  - Why are some people more likely to get a disease then others?

Terminology: allele

- ***Allele***: A specific variant at a location
  - The notion of alleles predates the concept of gene, and DNA.
  - Initially, alleles referred to variants that described a measurable trait (round/wrinkled seed)

  - Now, an allele might be a nucleotide on a chromosome, with no measurable phenotype.
  - As we discuss source of variation, we will have different kinds of alleles.

Terminology

- ***Locus***: The location of the ***allele***
  - A nucleotide position.
  - A genetic marker
  - A gene
  - A chromosomal segment

Terminology

- ***Genotype***: genetic makeup of (or part of) an individual
- ***Phenotype***: A measurable trait in an organism, often the consequence of a genetic variation
- Humans are ***diploid***, they have 2 copies of each chromosome, and 2 alleles at each locus
  - They may have ***heterozygosity/homozygosity*** at a location
  - Other organisms (plants) have higher forms of ***ploidy***.
  - Additionally, some sites might have 2 allelic forms, or even many allelic forms.
- ***Haplotype***: genetic makeup that is inherited or to be transferred as a package, e.g. (part of) a single chromosome

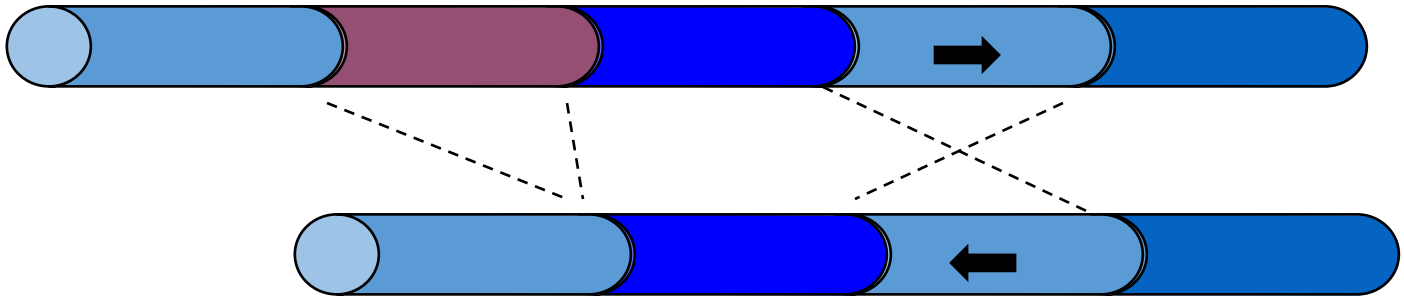
What causes variation in a population?

- Mutations (may lead to SNPs; insertions, deletions)

STRUCTURAL VARIATIONS

- Recombinations (may lead to insertions, deletions)
- Other crossover events e.g. gene conversion (may lead to increased or decreased copy numbers, insertions, deletions)

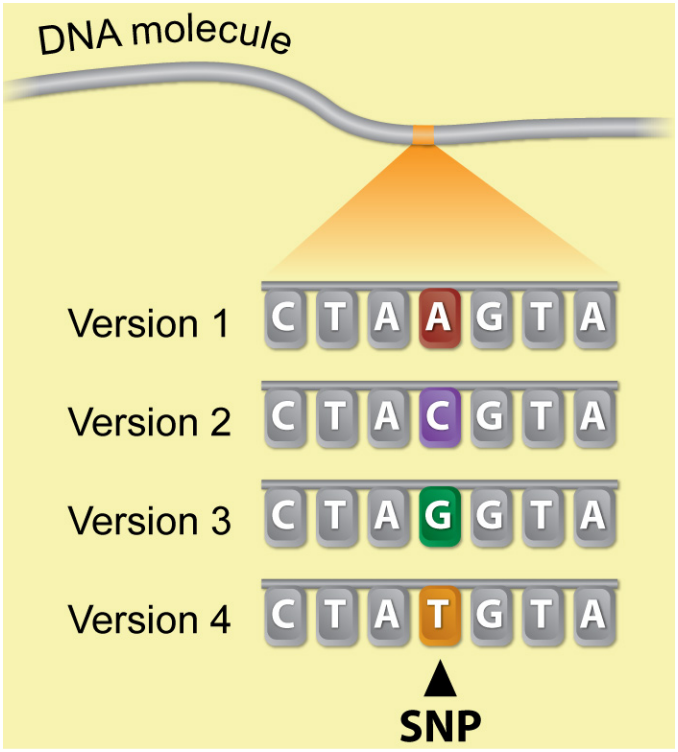
Structural polymorphisms



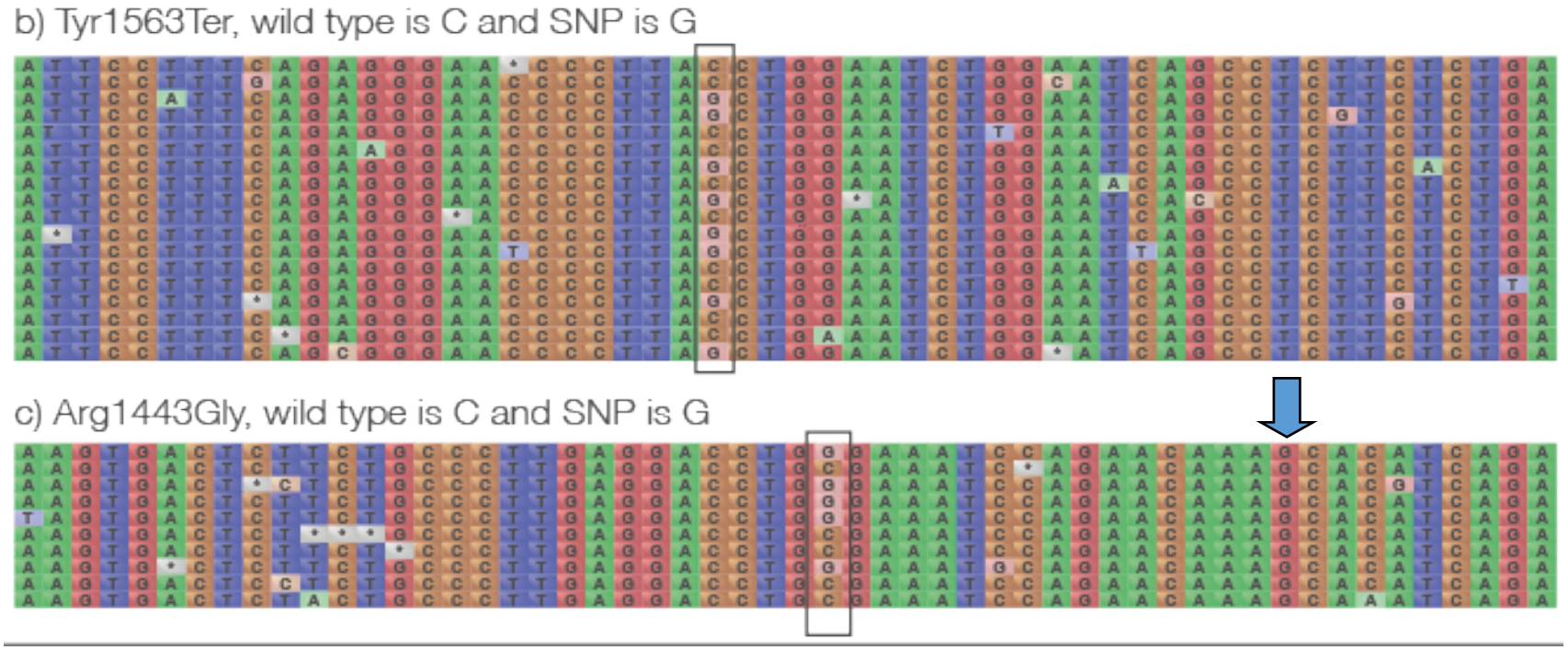
- Large scale structural changes (deletions/insertions/inversions) may occur in a population.
- Copy Number variation
- Certain diseases (cancers) are marked by an abundance of these events

Single Nucleotide Polymorphisms

- Point mutations that are sustained in a population are called **SNPs**
- SNPs are the **most common source** of variation studied



Single Nucleotide Polymorphisms



Infinite Sites Assumption:
Each site mutates at most once

The data is a matrix (rows are individuals, columns are loci). Only the variant positions are kept.

Copy number variations

Variable Number of Tandem Repeats (VNTR)

Minisatellites: 10-60 bp long canonic sequences, often repeated 5-50 times

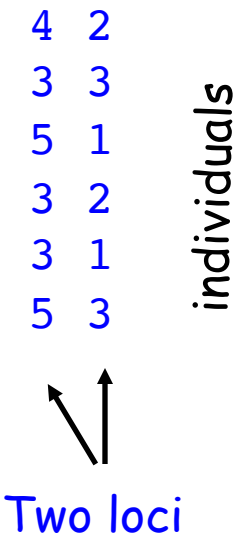
Microsatellites = Short Tandem Repeats (STR): 2-6 bp

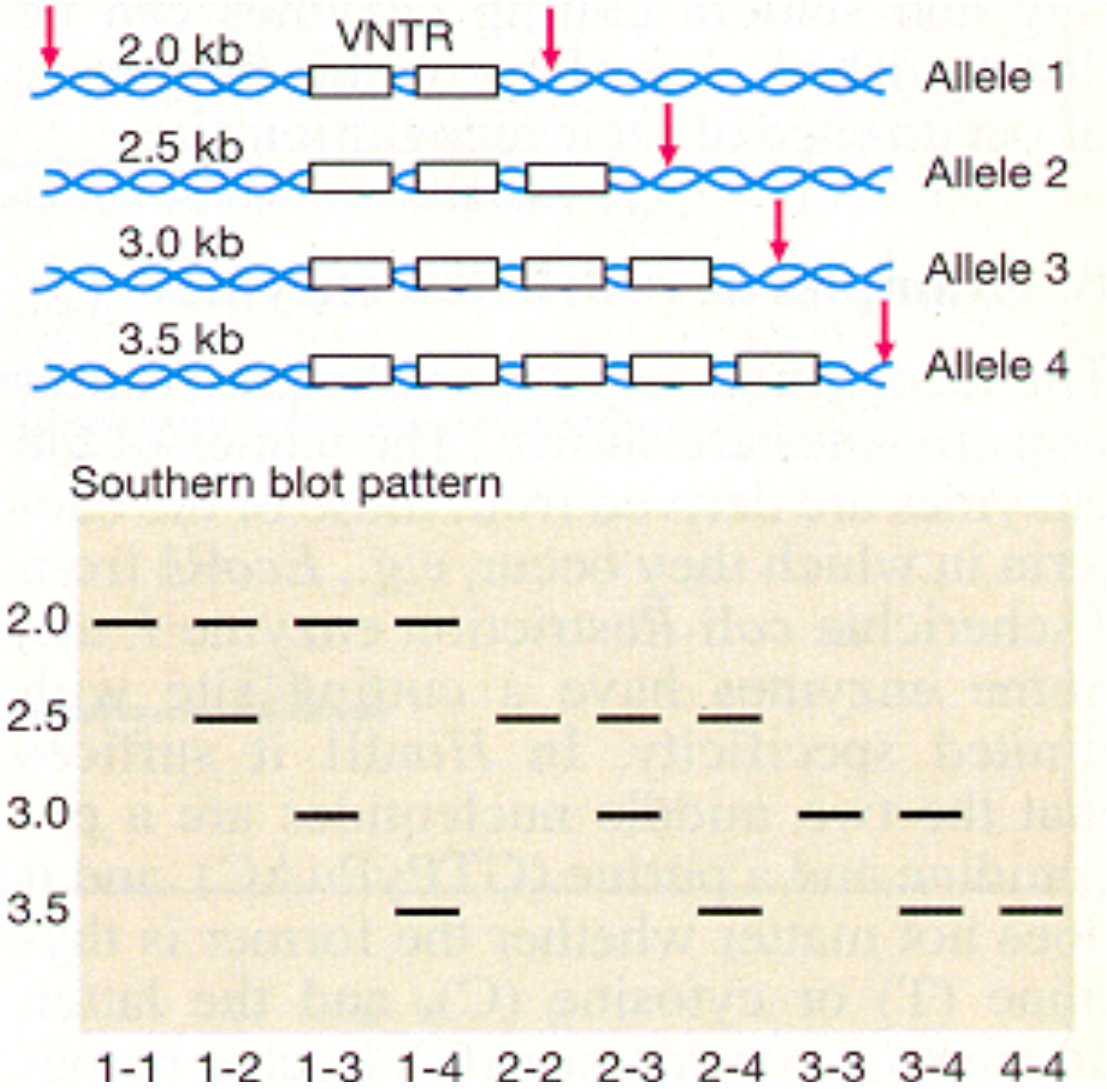
	N of repeats
GCTAGATCATCATCATCATTGCTAG	4
GCTAGATCATCATCATTGCTAGTTA	3
GCTAGATCATCATCATCATCATTGC	5
GCTAGATCATCATCATTGCTAGTTA	3
GCTAGATCATCATCATTGCTAGTTA	3
GCTAGATCATCATCATCATCATTGC	5

6 individuals compared for a certain locus that contains ATC repeats

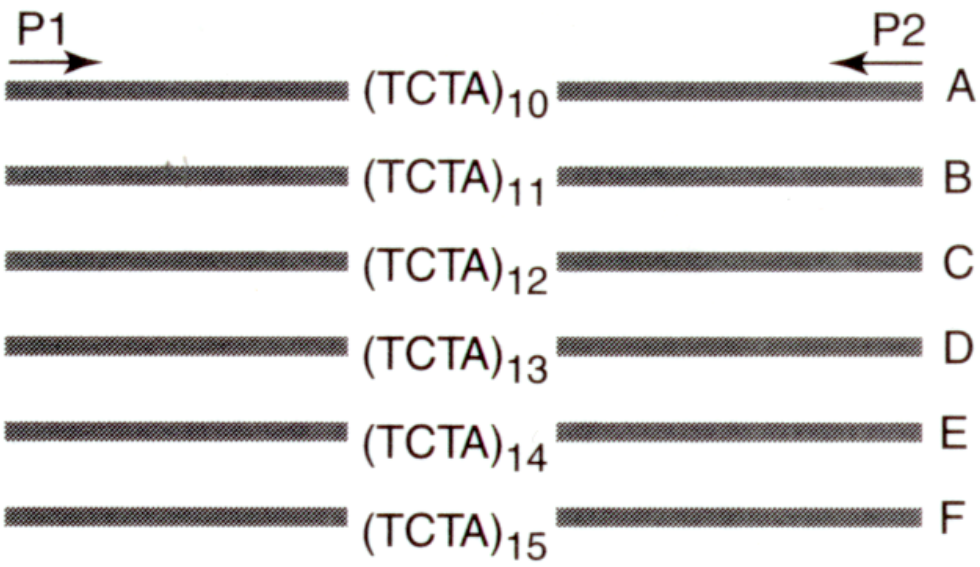
STR (VNTR) loci can be used as a DNA fingerprint

- Consider a collection of regions with VNTRs.
- Variable VNTRs will lead to variable length DNA
- The locations are far enough apart not to be linked
- The more the loci examined the more the combinational pattern gets personel

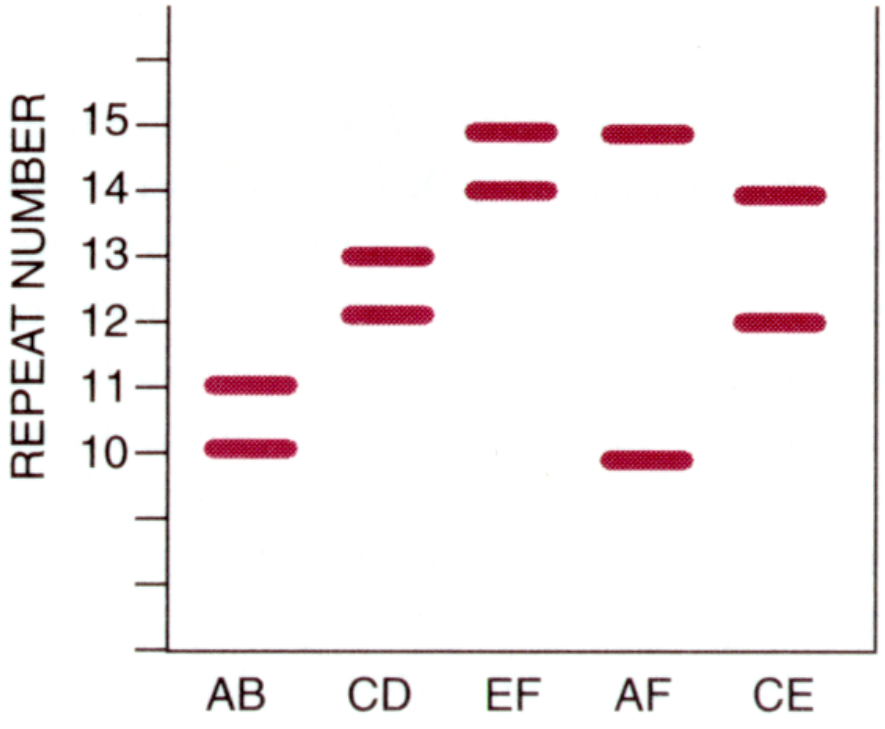


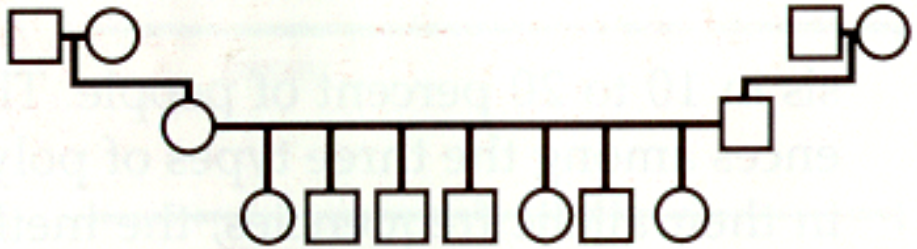


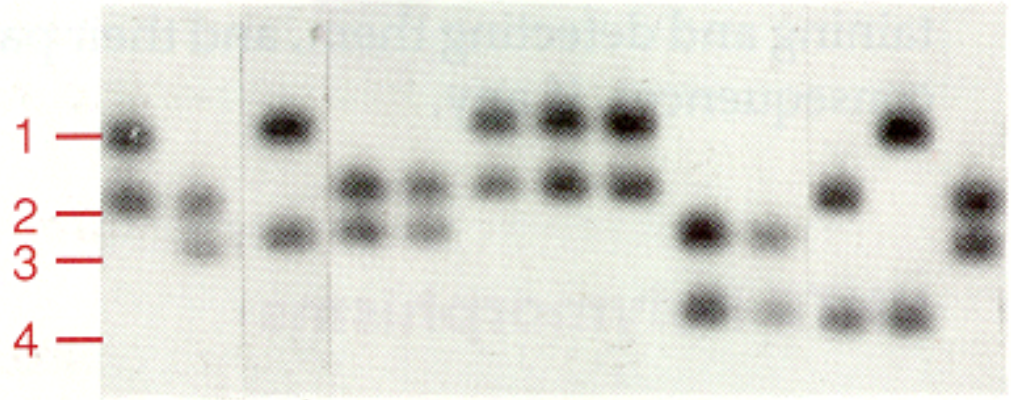
STR (VNTR) Polimorphism

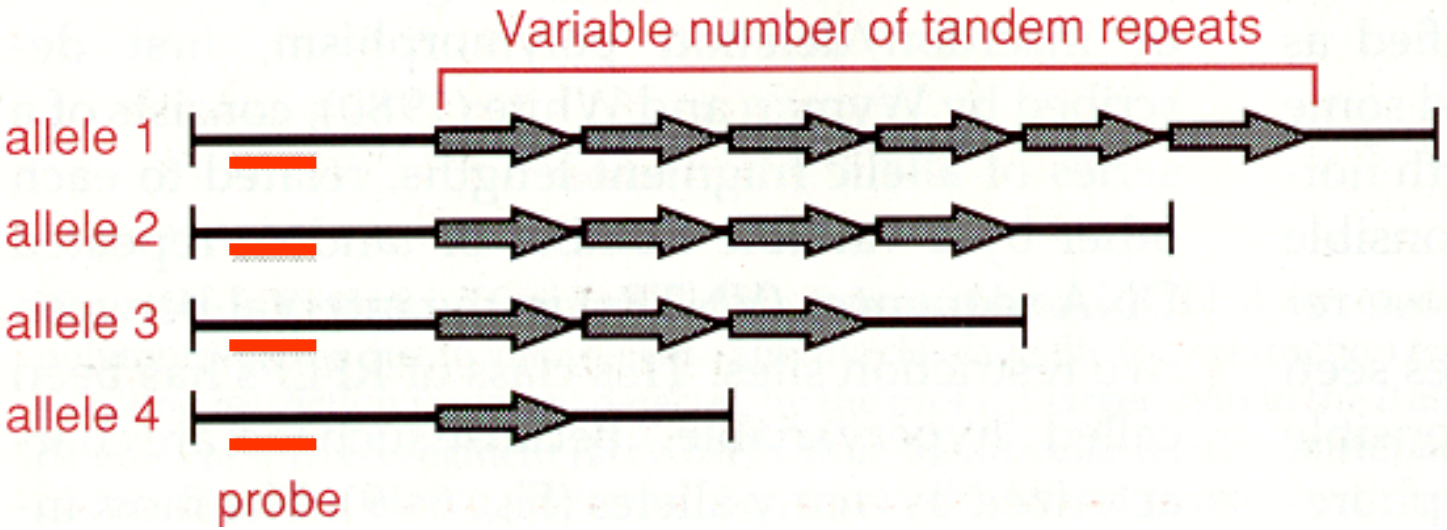


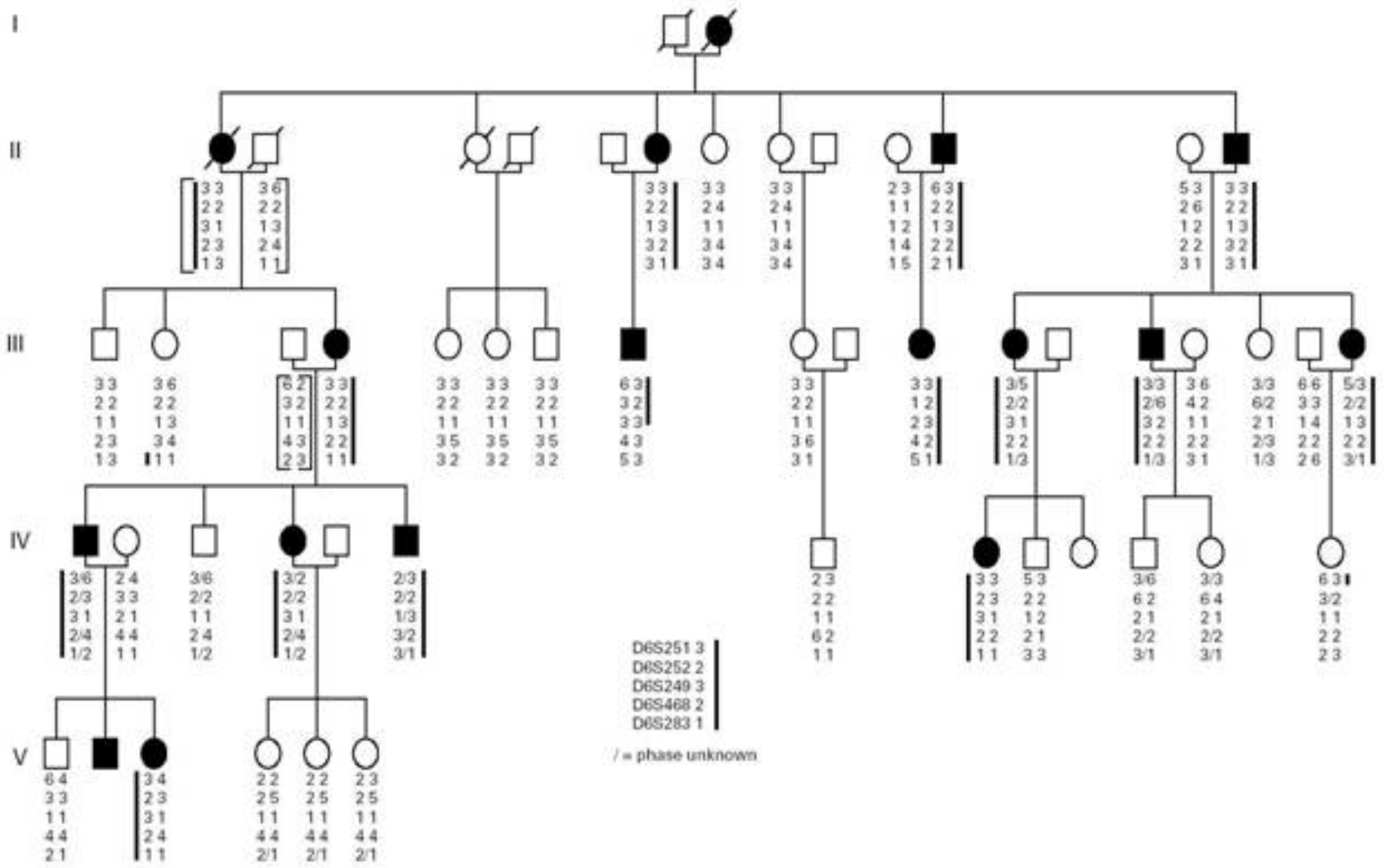
STR loci with 6 alleles



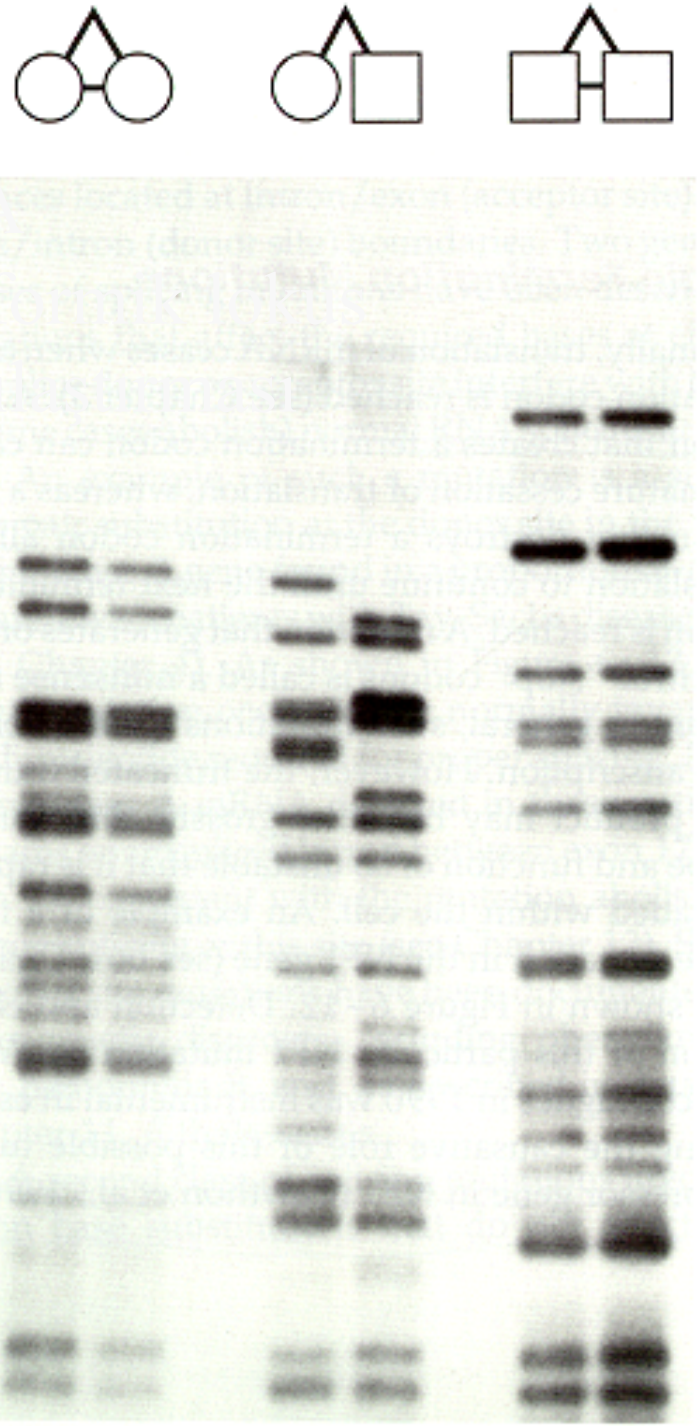



1
2
3
4


Variable number of tandem repeats
allele 1
allele 2
allele 3
allele 4
probe

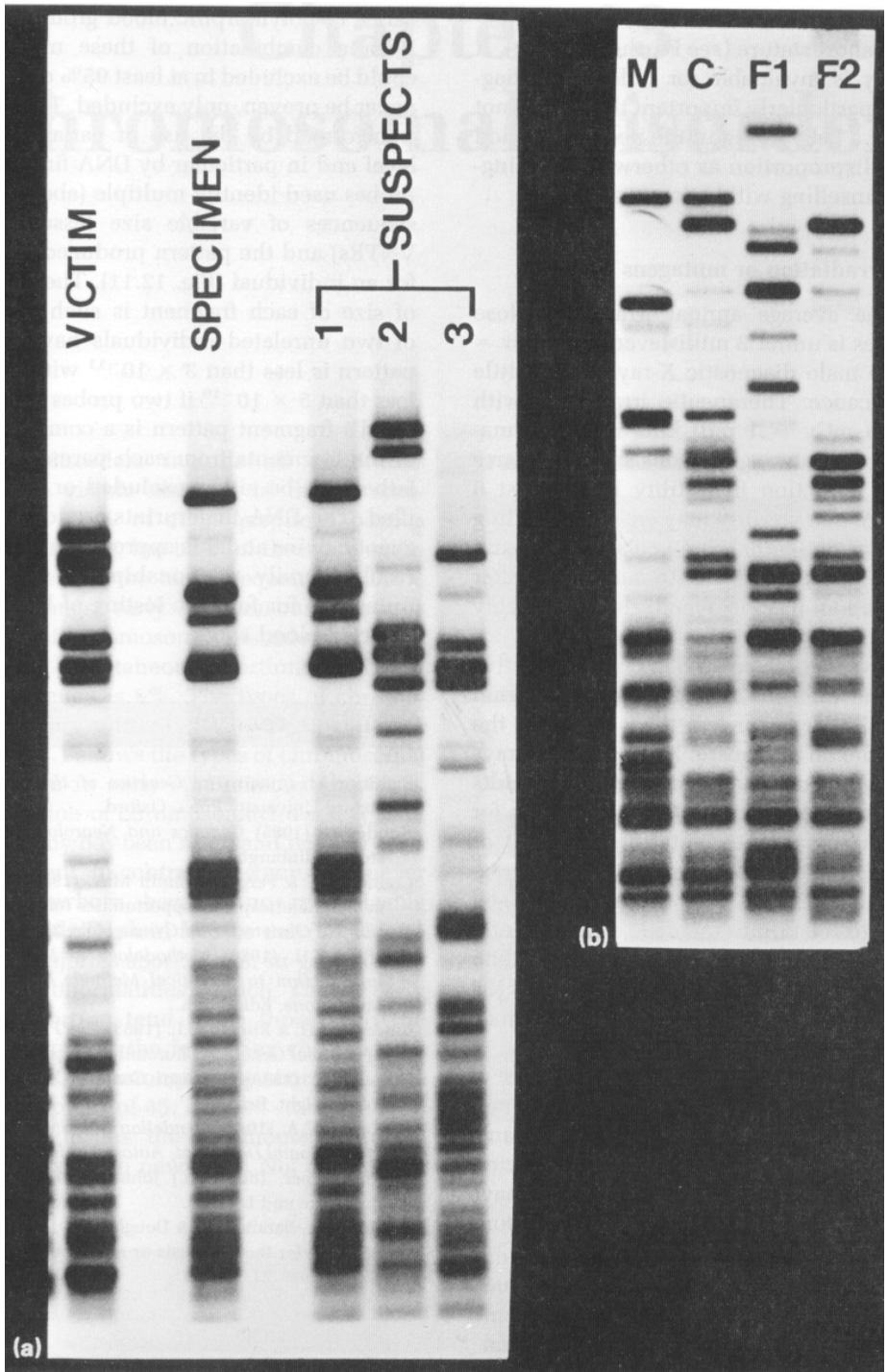


CNVs are transferred in Mendelian fashion unless there exist recombinations



Three pairs of siblings compared for
Their DNA fingerprints

DNA Fingerprint analysis in a criminal case



A forensic test laboratory uses a set of 13 different microsatellite markers in forensic analysis.
13 sets of specific PCR primers are used to determine the allele present in the test sample for each marker.
The marker used, the number of alleles at each marker and the probability of obtaining a random match for a marker is shown.

How often would you expect an individual to be mis-identified if all 13 markers are analyzed

Locus	No. of alleles random match	probability of
A	11	0.112
B	19	0.036
C	7	0.081
D	7	0.195
E	10	0.062
F	10	0.075
G	10	0.158
H	11	0.065
I	10	0.067
J	8	0.085
K	8	0.089
L	15	0.028
M	20	0.039

Probability= 0.112x0.036x0.081x0.195x----- =
= 1.7×10^{-15}
This joint probability that can be obtained from a set of polymorphic loci is called Polymorhism Information Content **(PIC)**

RFLP= Restriction fragment length polymorphism

Refers to variation (presence or absence) in restriction sites between individuals
Because of mutations in Restriction sites

These are extremely useful and valuable for geneticists

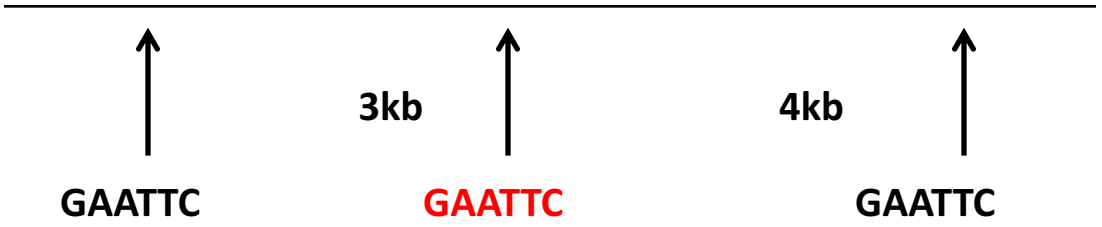
On average two individuals (humans) vary at 1bp in every 300-1000 bp

The human genome is $3x10^9$ bp

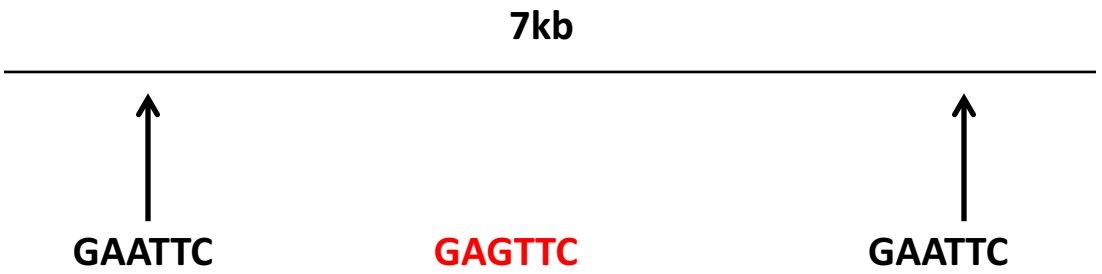
This means that they will differ in more than 3 million bp!!!

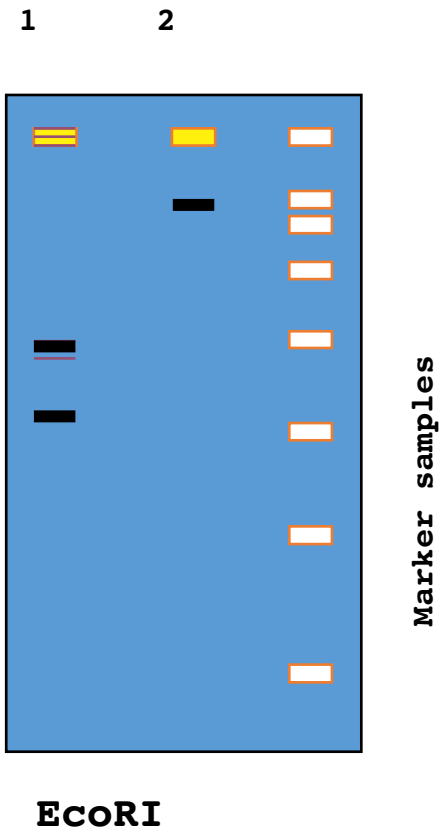
By chance these changes will create or destroy the recognition sites for restriction enzymes

EcoRI map for the region in one individual



Same region of a second individual may appear as

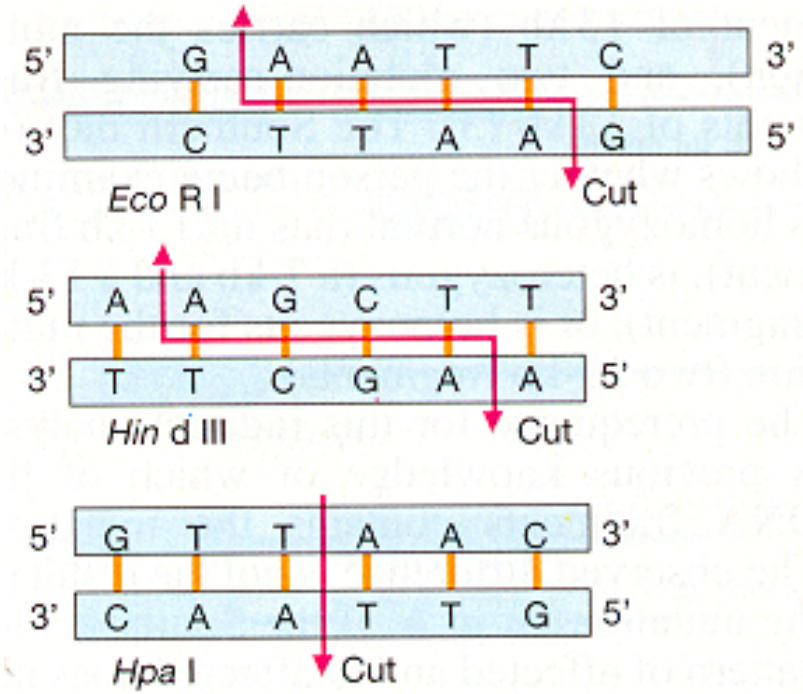




Variation#1 GAATTC
Variation#2 GAGTTC

This polymorphisms may or may not have any phenotypic consequences. Usually not!

Three restriction enzymes and their specific recognition sites



Example for some Polymorphic Serum Protein Groups

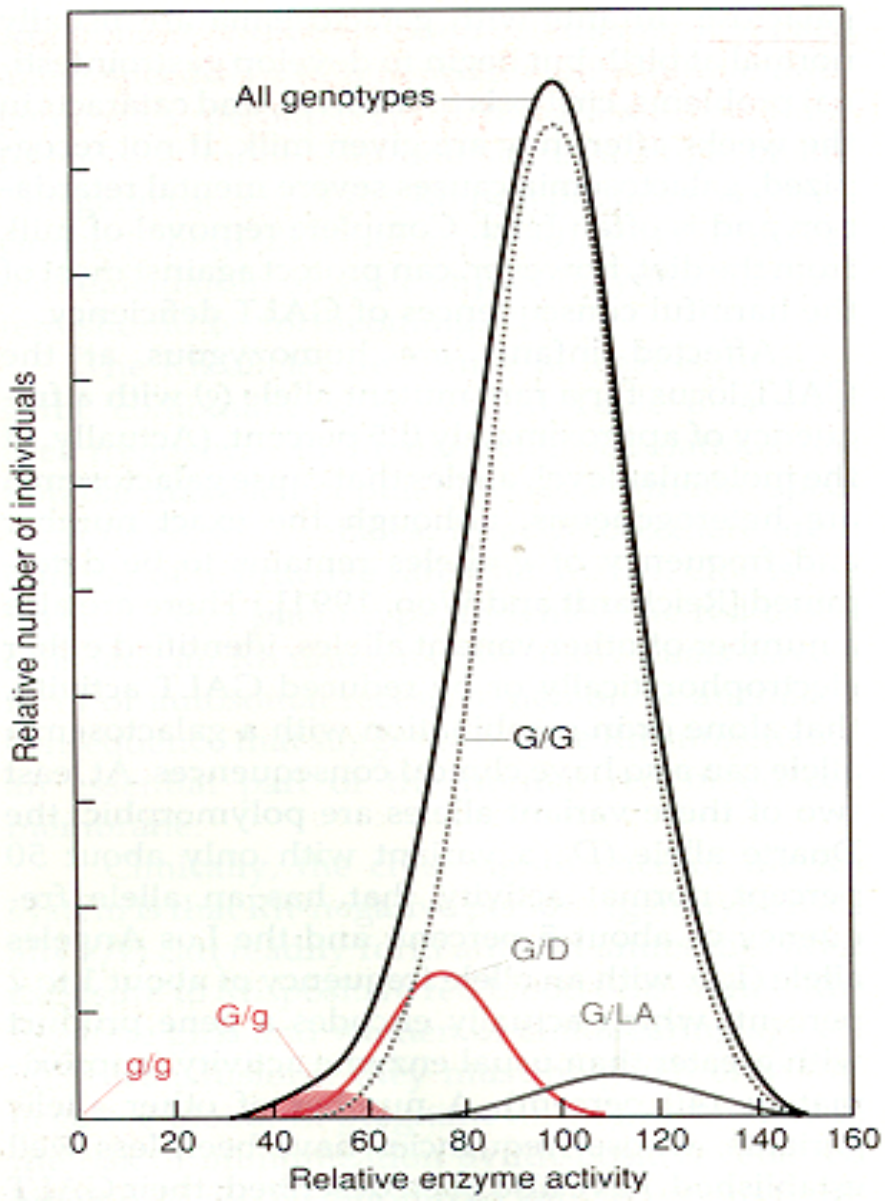
a,-antitrypsin	PI^{M1}, PI^{M2}, PI^{M3},
(a,-Protease inhibitor)	PI^{S}, PI^{z}
Ceruloplasmin	CP^{B}, CP^{A}, CP^{C}
Complement component-3	$C3^{s}$, $C3^{F}$
Group-specific protein	GC^{1F}, GC^{1s}, GC^{2}
Haptoglobin	HP^{1s}, HP^{1F}, HP^{2}
Immunoglobulins	Glm^{3}, $G3m^{5}$, $G1m^{1}$, $G1m^{12}$
IGHG (gm)	
IGKC (Km)	Km^{1}, Km^{3}
Properdin factor B	BF^{S}, BF^{F}
	(glycine-rich-b3glycoprotein)
Transferrin	TF^{C1}, TF^{C2}, TF^{C3}, TF^{B}, TF^{D}

Same enzyme may vary for its activity from individual to individual

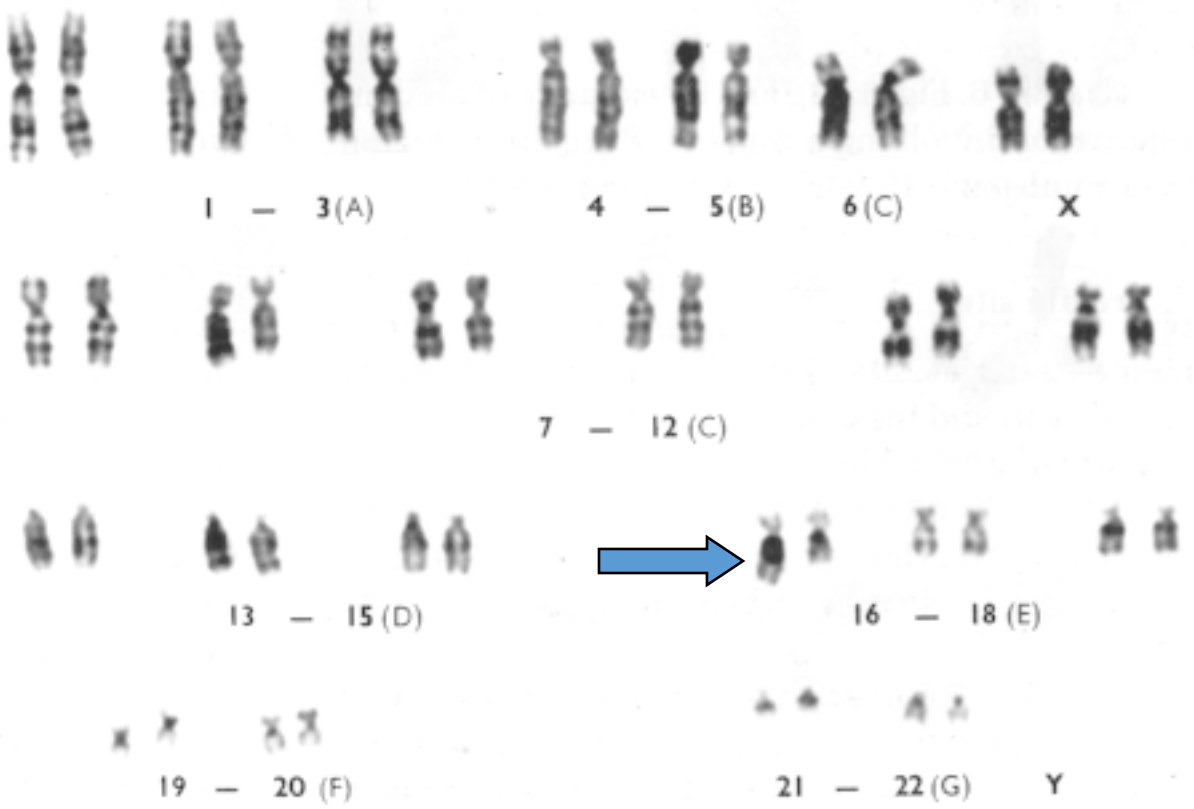
Which can be a normal variation (most usual)

Can be the cause of a disease phenotype

A consequence of its gene sequence



Polymorphic sites of chromosomes
Due to heterochromatin regions (1; 9; 16)



These are called chromosomal **heteromorphisms**

Evolution and Medicine

- Evolution of the Genomes

- Genetic Diversity and Polymorphism

- Evolutionary Mechanisms
  - **Microevolution vs Macroevolution**
  - **Population Genetics**
  - **Evolution and Medicine**

What is evolution?

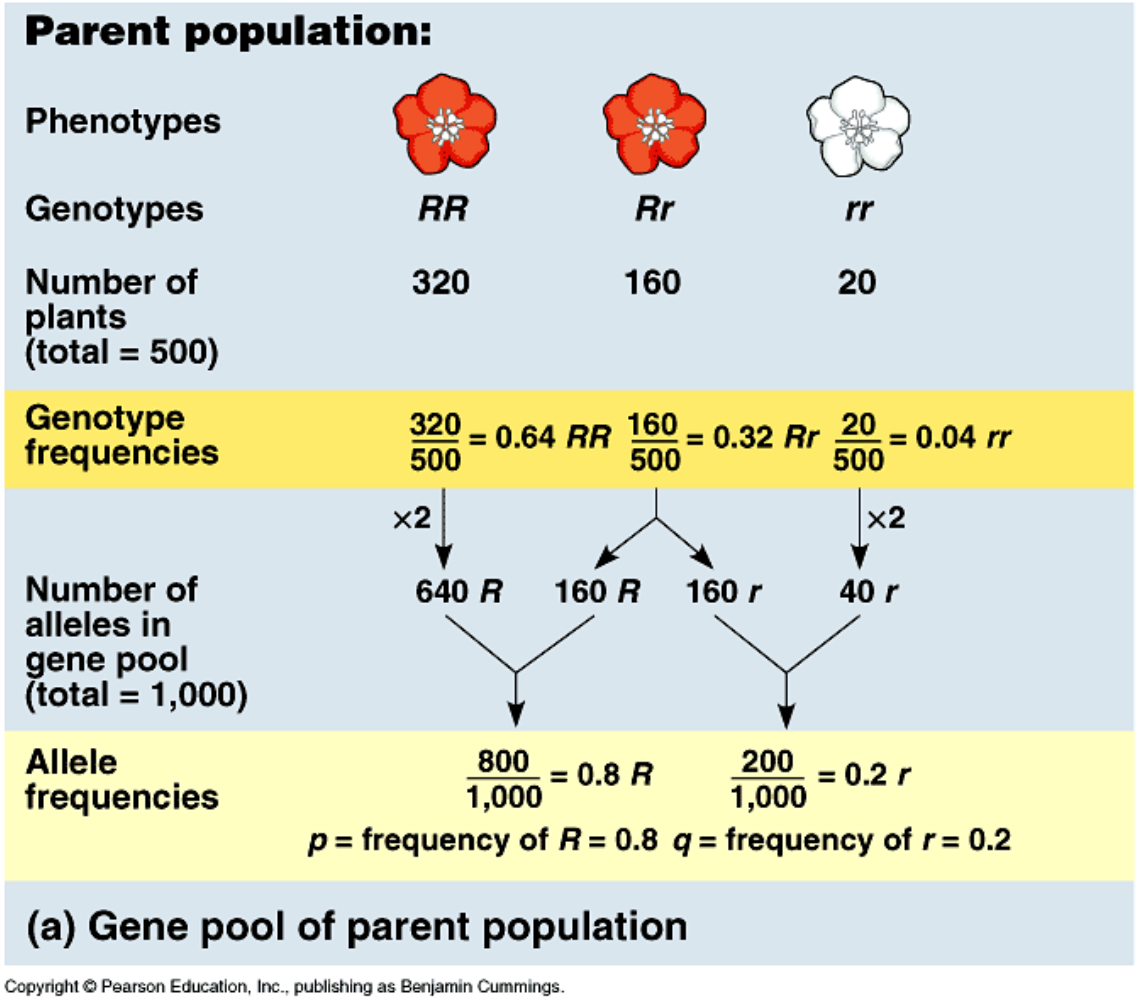
The change in the genetic make-up of a species over time

Change of the allele frequencies in the gene pool of a population over time

POPULATION GENETICS

- Study of evolution from a genetic point of view
- What is a population?
  - A collection of individuals of the same species that routinely interbreed (living in the same location)
- A species is a naturally breeding group of organisms that produce fertile offspring.
- Evolution occurs in the population

  (not in the individual)



Gene pool – total genetic information available in a population.

Allele frequency tends to remain the same from generation to generation unless acted upon by an outside force.

Hardy-Weinberg Equilibrium

- The allele frequency in a population can be calculated using the Hardy-Weinberg equation.
- This equation states that all of the allele combination must add to ONE (1)
- Dominate alleles + recessive alleles = 1
- $p + q = 1$
- The frequency of the heterozygote may also be calcuated

What is Hardy-Weinberg Equilibrium?

- Assumes NO evolution occurs.
- All 5 conditions must be meet.
- Can never happen!
- It is a model or a yardstick to measure how much a population or species has evolved.

Hardy-Weinberg Equilibrium Conditions

- No net mutation
- No migration
- Large population size (best infinite)
- Mating is random
- Natural selection does not occur

Disruption of Genetic Equilibrium

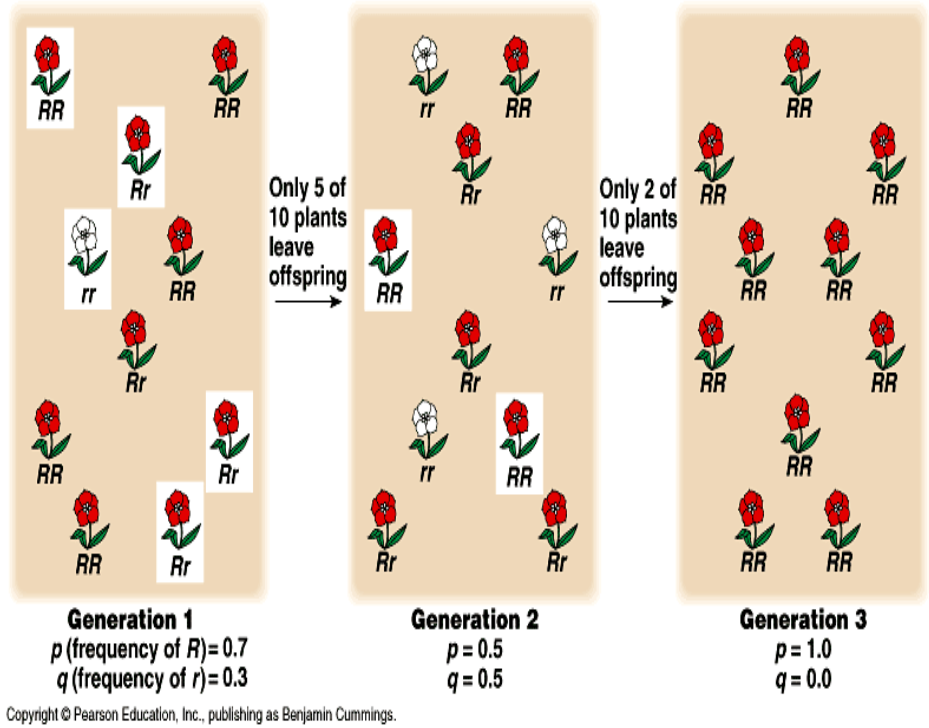
- Mutations (are random, increase variation)
- Migration in/out (effects population size as well)
- Genetic Drift
- Non-random mating
- Natural Selection

MIGRATION

- Immigration
- Emigration
- Gene flow – moving genes from population to another

Genetic Drift (is a random process, decreases variation)

- Change in the allele frequency as a result of random events or chance
  - Usually occurs in small populations
  - After a natural disasters
    - Flood
    - Forest fire



In the smallest population allele frequency reaches 0 after the 45th generation = no variation

Non-random Mating

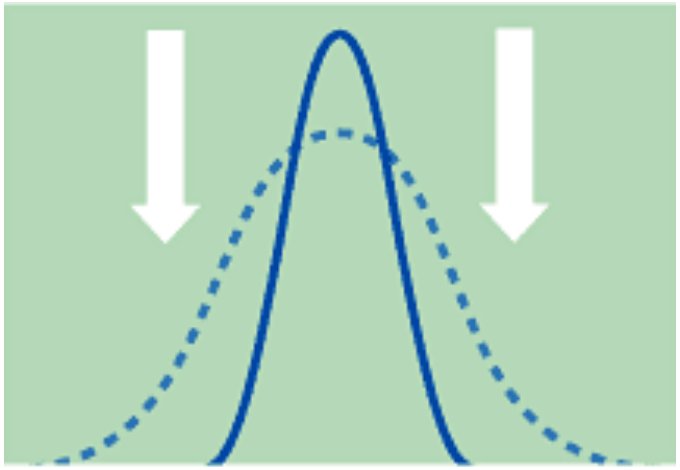
- Random Mating – mating without regard to genetic make-up
- Sometimes mating selection is often influenced by geographic proximity
- Many animals do not mate randomly

NATURAL SELECTION

1. All species have genetic variation.
2. The environment presents many challenges
3. Organisms tend to produce more offspring than the environment can support - competition (struggle for survival)
4. Some individuals are better suited to cope with the challenges (survival of fittest)
5. Characteristics best suited to environment tend to increase in a population over time
6. Natural selection **decreases variation**
7. Natural selection **is non-random**
8. Three main selection types

STABILIZING SELECTION

- Individuals with the average form are of a trait have the highest fitness



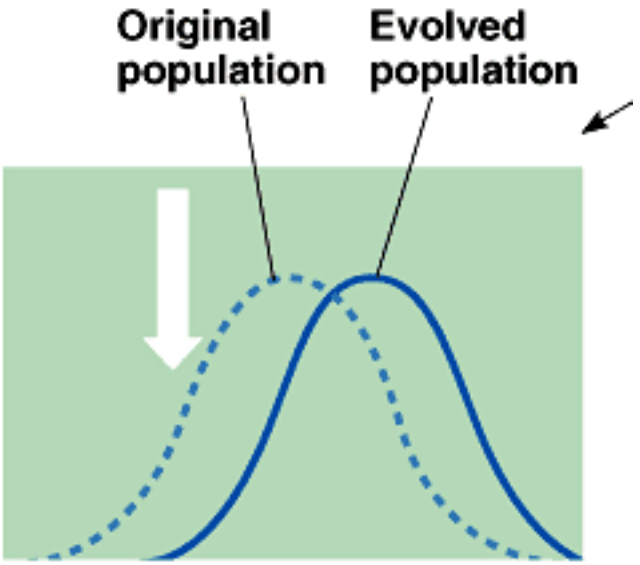
Example:

1. Birth weight, head circumference in offspring
2. Seed size

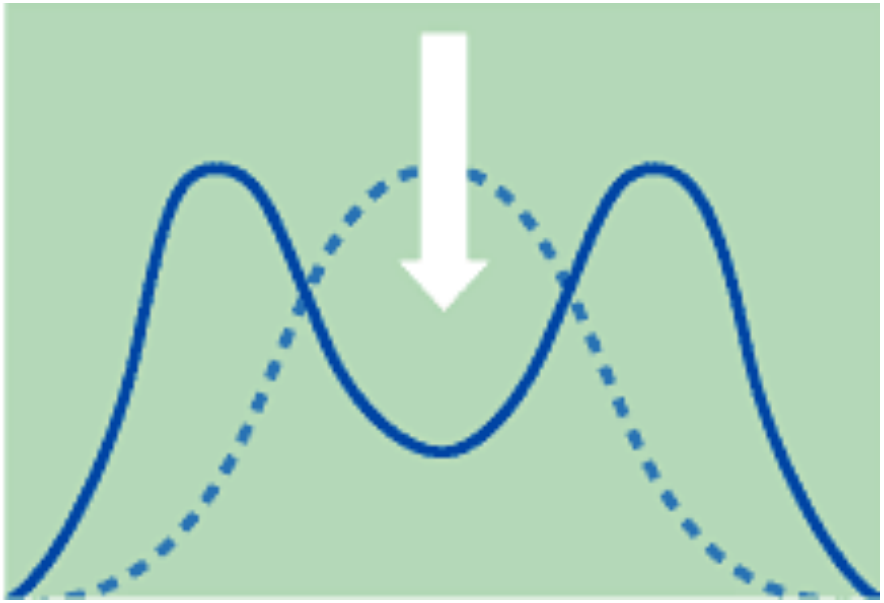
DIRECTIONAL SELECTION

- The frequency of one allele tends to move in one direction (more of one of the extremes forms of the trait

Example – tongue length in anteaters



DISRUPTIVE SELECTION



- Individuals with either extreme have an advantage over individuals with the average form of the trait.
- Example: Limpet shell coloration

Sexual Selection

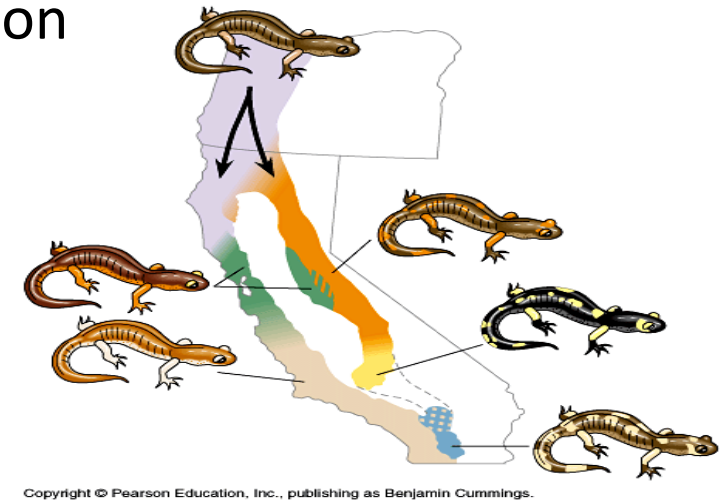
MICROEVOLUTION
LEADS TO
MACROEVOLUTION

SPECIES FORMATION

- Begins with geographic isolation
- Results in reproductive isolation
- Speciation occurs

MECHANISM FOR REPRODUCTIVE ISOLATION

- Geographical isolation
- Ecological isolation
- Temporal isolation
- Mechanical isolation
- Reproductive failure



How does macroevolution/speciation occur?