Biology or Genes?

- Biological variation
  - Genetics
    - This is what we think of when we say biological differences
      - Race implies genetics
  - Physiology
    - Not all physiological variation is genetically mediated
      - Tanning, Sweating, etc.
  - Development
    - Much difference in brain function is based on neuronal connections that are environmentally mediated, especially up to 3 years of age

## How about the genes?

- Individual
  - DNA sequence information
    - About 3 billion nucleotides
    - Determines protein structure, timing and magnitude of protein synthesis
- Population
  - Gene frequencies influenced by mating patterns, mutation, migration, random factors (genetic drift, founders effect), and natural selection

DNA Structure

- Ladder or lattice-like molecule composed of smaller building blocks: **Nucleotides**
- Nucleotides are composed of three subunits:
  - Sugar (deoxyribose for DNA)
  - Phosphate
  - Nitrogenous base--the key to the functioning of DNA
    - Purines (2 N/C rings): Adenine, Guanine
    - Pyrimidines (1 N/C ring): Thymine, Cytosine

DNA

- The sides of the DNA "ladder" are composed of bonded sugar and phosphate, the "rungs" are based on hydrogen bonds between complementary nitrogenous bases
- The key to DNA function is the complementarity of the bases: Adenine bonds only with Thymine and Guanine only with Cytosine.
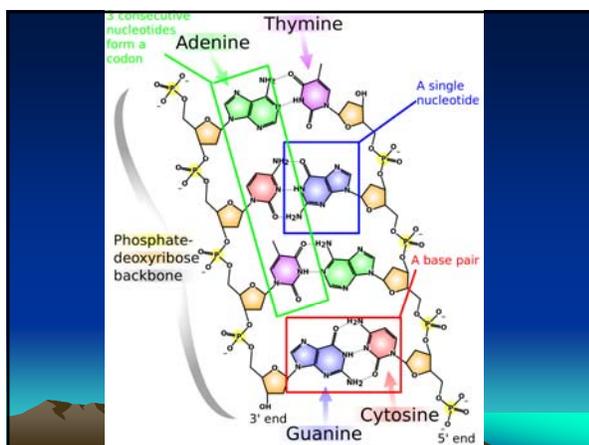
Proteins

- Life functions are regulated by proteins:
  - **Structural proteins** like the muscle cell building block, myosin
  - **Regulating proteins**, including enzymes and hormones, like the enzyme that breaks down starch molecules, salivary amylase
 - **Transporter proteins** like the oxygen transporter, hemoglobin

Protein Structure and Function

- Function of different proteins is based on structure
- Structure determined by the number and type of building blocks, called Amino Acids
- Amino acids are assembled into chains called polypeptides
- A functional protein may include several polypeptides

Protein Synthesis

- Sequence of amino acids is determined by sequence of bases in the DNA unit (or gene)
- Protein synthesis is a two-step process:
  - **Transcription**: copying the DNA to mRNA
  - **Translation**: using mRNA to assemble the polypeptide with tRNA carrying amino acids
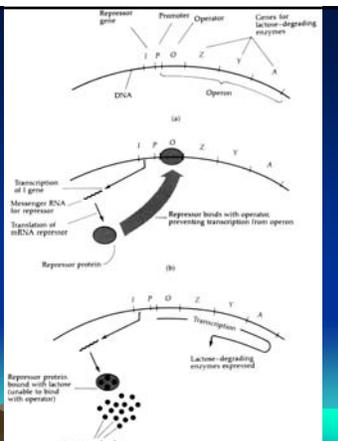

DNA to RNA Transcription

Translation

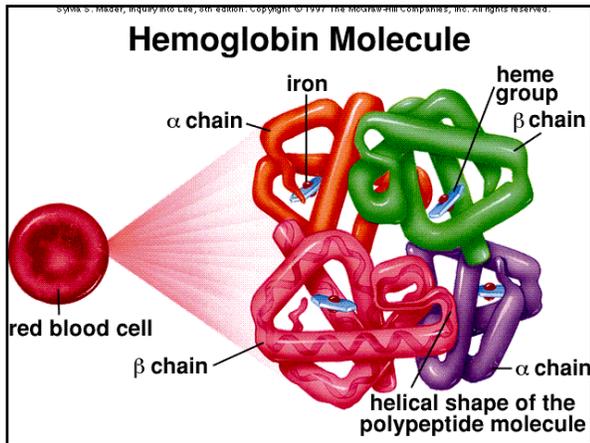5' cap | AUGAGAUACCAAGAACCUACCAAGGUAGAGCUUUAGCCCG | AAAAAAAAAAAAA | 3'
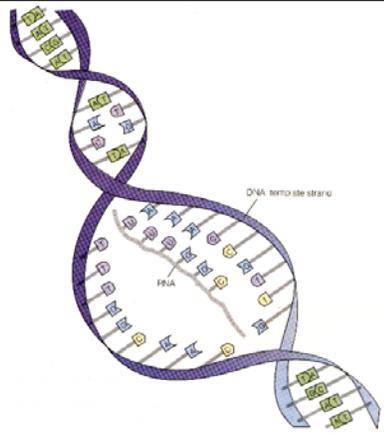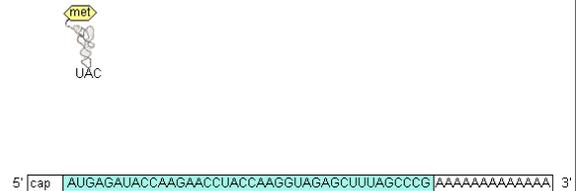
Gene Regulation

- Humans and chimps share ~98% of their genes, but the regulation of those genes is what makes for the physical differences
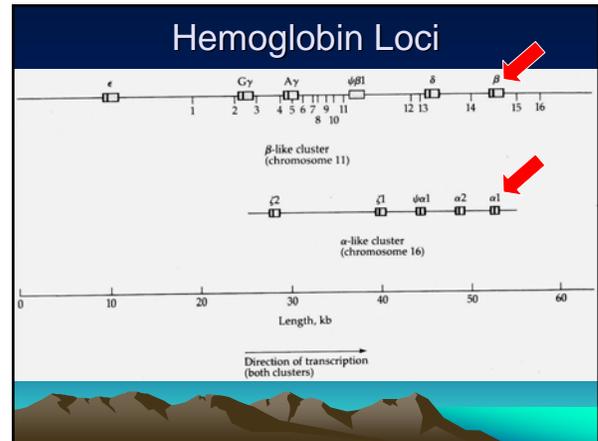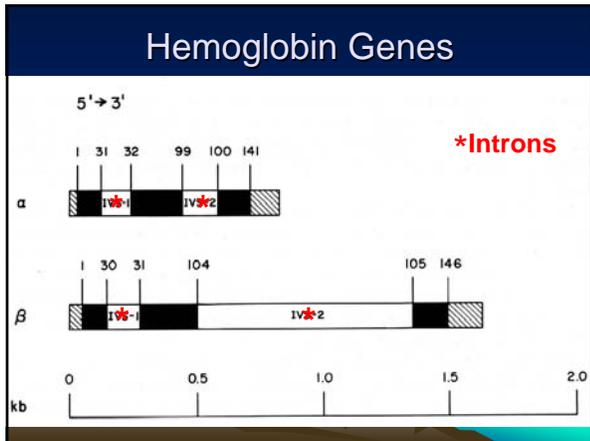

How does this work in Humans?

- The Hemoglobin molecule is a complex protein structure that carries oxygen and carbon dioxide through the blood stream
  - It consists of four polypeptides: 2 alpha and 2 beta chains
  - Each of these polypeptides has a separate section of DNA carrying the code for the appropriate sequence of amino acids

Hemoglobin Molecule

- iron
- α chain
- heme group
- β chain
- red blood cell
- β chain
- α chain
- helical shape of the polypeptide molecule

Hemoglobin

- Each alpha chain consists of 141 Amino Acids, requiring a sequence of 423 nucleotides in the DNA of the 16th Chromosome
- Each beta chain consists of 146 Amino Acids, requiring a sequence of 438 nucleotides in the DNA of the 11th Chromosome

---

Hemoglobin Genes

$5' \rightarrow 3'$

*Introns

α   I 31 32   99 100 141
     IVS-1   IVS-2

β   I 30 31   104   105 146
     IVS-1   IVS-2

kb   0   0.5   1.0   1.5   2.0

Hemoglobin Loci

ε   Gγ Aγ ψβ1   δ   β
     1   2 3 4 5 6 7 9 11 12 13 14 15 16
                  8 10

β-like cluster
(chromosome 11)

ζ2   ζ1 ψα1 α2 α1

α-like cluster
(chromosome 16)

0   10   20   30   40   50   60
Length, kb

Direction of transcription
(both clusters)

Chromosomes

- Chromosomes are the complex DNA and Protein units that carry the genetic code in all cells with nuclei
- In sexually-reproducing organisms, chromosomes come in homologous pairs
  - Each member of the pair contains information on how to build the same protein products
  - One member of each pair comes from the mother and one comes from the father

---

- DNA
- Gene
- Gene
- Proteins
- DNA and proteins are multi-folded
- Chromosomes
- Homologous chromosomes
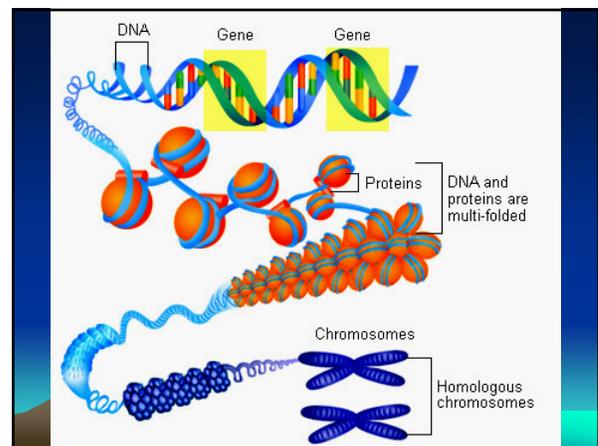
Human Genome

- Humans have 23 pairs of chromosomes
  - 22 pairs of **autosomal chromosomes** affecting almost all aspects of the individual other than sex
  - 1 set of **sex chromosomes**
    - A pair of X chromosomes for Females
    - One X and one Y chromosome for Males
- Approximately 25,000 genetic loci on the 23 pairs of chromosomes

Locus

- The position of a gene on an homologous chromosome pair is known as a **Locus**
  - The locus of the betaglobin gene for the Hemoglobin molecule is near the tip of the short arm of chromosome number 11
  - The locus of the alphaglobin gene is near the centromere (pinching near the middle of the chromosome) on the short arm of chromosome number 16
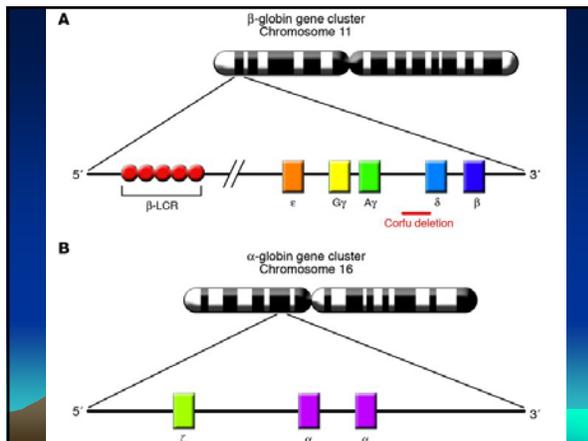
Alleles

- Many genes have different forms
- The ABO blood group has three alleles
  - Each allele causes a different substance to be attached to the outer wall of red blood cells
  - A, B, O molecules differ slightly in structure
- These variants of a particular gene are called **Alleles**
- A **polymorphism** (literally many forms) is a genetic locus with two or more alleles that occur in appreciable (>1%) frequency in a given population

Genotype vs. Phenotype

- **Genotype** is the genetic makeup of an individual
  - Usually refers to what pair of alleles an individual has at a specific locus on a given chromosome pair
- At the ABO locus (9th Chromosome), an A allele from the mother and an O allele from the father = genotype AO
  - There are 6 possible genotypes based on the three ABO alleles
    - AA, AO, BB, BO, AB, OO

Genotype vs. Phenotype

- **Phenotype** is the observable expression of the genotype
  - The phenotype for the AO genotype would be Blood Type A
    - There are 4 possible phenotypes for ABO
      - Blood Type A (AA, AO), Blood Type B (BB, BO), Blood Type AB, Blood Type O (OO)
- Distinction made more sense before DNA sequencing was possible

Homozygous

- If an individual has two of the same alleles at a particular locus, he is said to be **homozygous (or is a homozygote)**
  - A person is homozygous if he inherits a Hemoglobin S allele from both his mother and father
  - Genotype: **Hb^S/ Hb^S or SS**
  - Phenotype: Sickle Cell Anemia
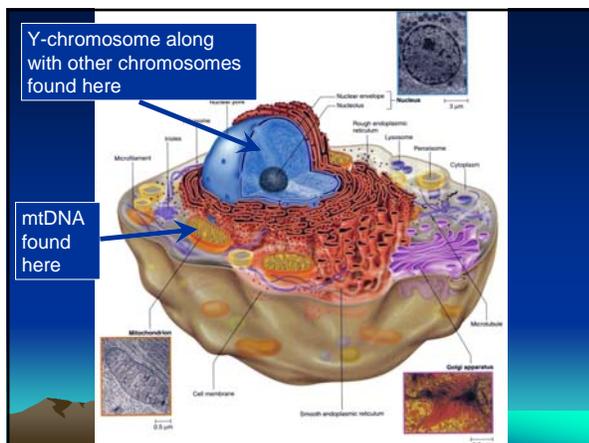
Heterozygous

- If an individual has two different alleles at a particular locus, he is said to be **heterozygous (or is a heterozygote)**
  - A person is heterozygous if he inherits a Hemoglobin S allele from his mother and a Hemoglobin A allele from his father
  - Genotype: **Hb^S/ Hb^A or SA**
  - Phenotype: Sickle Cell Trait (Carrier)
- We use heterozygosity as an indicator of population structure (more later)

Dominant and Recessive

- Alleles are said to be dominant or recessive depending upon whether they are expressed (**dominant**) or hidden (**recessive**) in heterozygotes
  - In the ABO system, A and B alleles are dominant over O, and **co-dominant** with each other (Blood type AB)
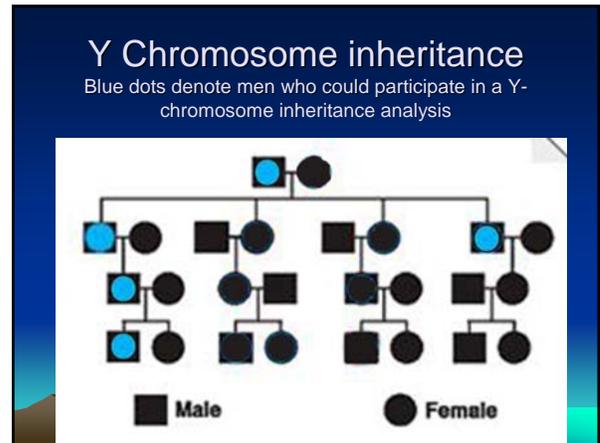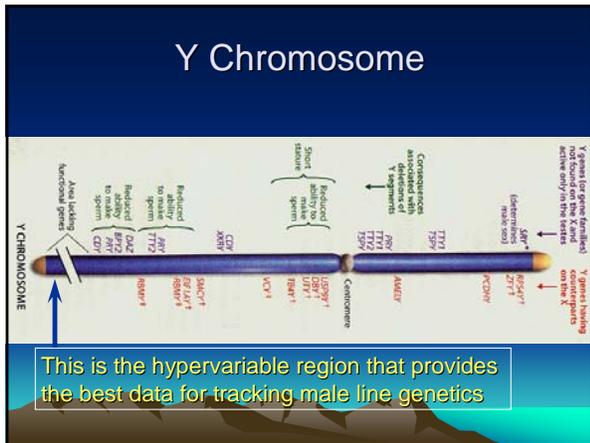 - O is recessive to both A and B

Predicting Offspring Types

GAMETES	Father A	Father O
Mother A	AA	AO
Mother O	AO	OO

Blood Type A 3 out of 4 = 75%

Blood Type O 1 out of 4=25%

Y-chromosome along with other chromosomes found here
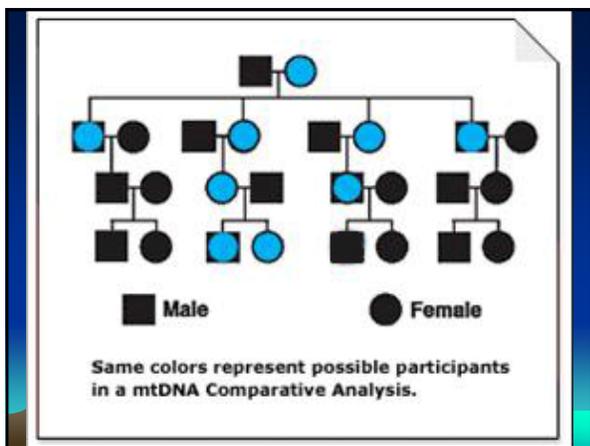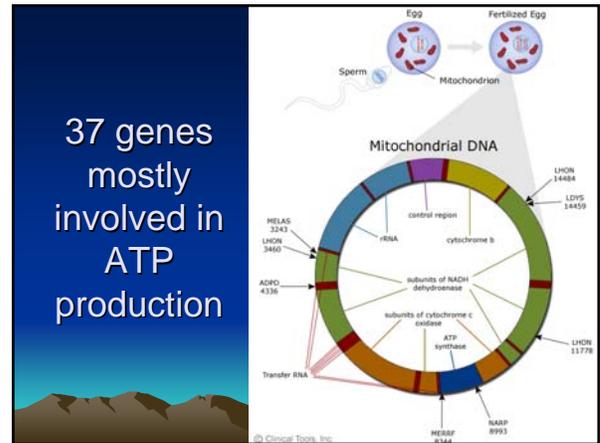
mtDNA found here

Y Chromosome

- The presence of a Y-Chromosome causes maleness
  - This little chromosome, about 2% of a father's genetic contribution to his sons, programs the early embryo to develop as a male
  - It is transmitted from fathers only to their sons
  - Most of the Y-Chromosome is inherited as an integral unit passed without alteration from father to sons, and to their sons, and so on, unaffected by exchange or any other influence of the X-Chromosome that came from the mother
    - It is the only nuclear chromosome that escapes the continual reshuffling of parental genes during the process of sex cell production.

Y Chromosome

This is the hypervariable region that provides the best data for tracking male line genetics

---

Y Chromosome inheritance
Blue dots denote men who could participate in a Y-chromosome inheritance analysis

■ Male   ● Female

The "other" DNA

- Mitochondrial DNA is inherited ONLY in the maternal line
  - All mitochondrial DNA in any human came from that individual's mother
  - It is not altered by sexual reproduction
  - Changes come only from mutations during cell division
  - Mutations that occur in the control region tend not to be repaired, since that region does not code for any specific product
  - Only mutations that occur in female germline cells (those that become eggs) are passed on to offspring
- Mitochondrial DNA is circular

---

37 genes mostly involved in ATP production

■ Male   ● Female

Same colors represent possible participants in a mtDNA Comparative Analysis.

SNPs
Pronounced "snips"

- Single Nucleotide Polymorphism
  - ~1/1,300 (<0.1%) of the nucleotides differ between any two people
  - example:
    John   TTGACGTCA**GT**GCCGTGAC
    Dave   TTGACGTCA**CT**GCCGTGAC

  - Contrast to chimpanzee: 1.4% of aligned bases are different (plus ~3% insertions and deletions of varying sizes)

SNPs

- Stable snips are relatively rare
- So infrequent they have occurred at a particular position in the genome only once in the course of human evolution
- Stable snips have been termed "**u**nique **e**vent **p**olymorphisms" (**UEP**s)

STRs

- Short Tandem Repeats (STRs) or Microsatellites are repeats of 2 to 6 bases (e.g., acg/acg/acg/acg/acg/acg…)
  - STRs are commonly used for forensic work
  - The number of repeats is counted and that number is the raw data for comparisons
  - The specific number of repeats in a particular variant (or allele) usually remains unchanged from generation to generation but changes do sometimes occur and the number of repeats may increase or decrease

STRs

- Changes in microsatellite count occur much more frequently than new UEPs arise
  - A UEP has arisen only once in humans, the number of repeat units in an STR may have changed many times along a paternal lineage
- STRs can facilitate the estimation of population divergence times
  - Which can then be compared with estimated mutational ages of the UEPs
  - The combination of these two kinds of data offers a powerful tool with which to assess patterns of migration, admixture, and ancestry

Minisatellites

- Minisatellites have a repeating motif of between 8 and 100 bases (e.g., acgcattagt/acgcattagt…)
  - These motifs are long enough that there tends to be variation in the sequences
  - For comparison purposes need to count repeats and enumerate major motif variations
  - Changes take place more frequently in minisatellites than in STRs

The Y-Chromosome Evolutionary Clock

- UEPs are the hour hand
- STRs are the minute hand
- Minisatellites are the sweep second hand
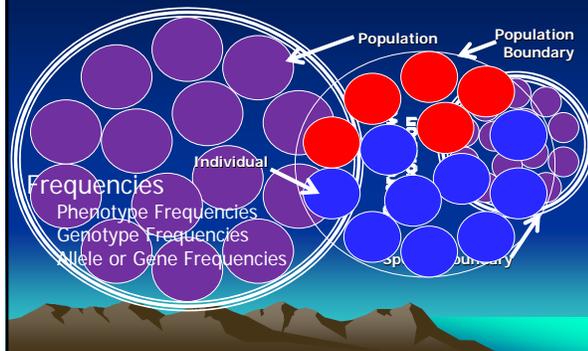

Types of Mutations
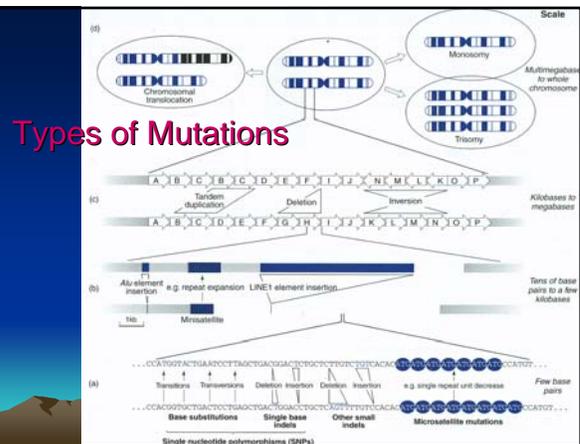
How do genes get distributed across the species?

They sort out by populations

- What is a population?
  - A population is a subdivision of a species
  - A population is a local community of individuals where mates are usually found
    - The likelihood of finding a mate within the population is much greater than finding a mate in another population
  - A population shares a common gene pool
  - A population has continuity through time

Population ≠ Race

Populations within a species

Population
Population Boundary
Individual
Frequencies
Phenotype Frequencies
Genotype Frequencies
Allele or Gene Frequencies

How do we know if people belong to the same population?

- They need geographical proximity
  - Must be able to meet to mate!
- Share a common language
  - Communicate with one another
- Share ethnicity, culture, religion
  - More likely to mate if sharing history and values

Estimated millions of human populations in the world at the start of the 21st Century

Where do we find our mates?

- Until recently most people stayed in the place they were born and married someone from nearby
- The last half of the 20th century was characterized by the breakups of local population isolates
- This major change in breeding pattern has resulted in increased genetic heterogeneity

Figure 12.3. The percentage of marriages that took place at various marital distances (the distance between the premarital residences of bride and groom) for the town of Barre, Massachusetts, 1800–1849. (From unpublished data of J. H. R.)

Mating Patterns

- Any non-random mating produces changes in genotype distribution
- Assortative Mating
  - Mating at greater than random frequencies between individual with like (Positive) or unlike (Negative) genotypes
    - Positive Assortative Mating (skin color, height, etc.) increases homozygotes
    - Negative Assortative Mating (immune system) increases heterozygotes

Mating Patterns

- Inbreeding is mating at greater than random frequencies between biologically related individuals
  - Incest taboos prohibit mating between closely related individuals, making inbreeding less common than random
    - Increases heterozygosity
  - Cousin marriage preferences, make inbreeding more common than random
    - Increases homozygosity (Polygamy and fumarase deficiency)

Evolution of Genes in Populations

- **Mutation** is the source of all new genetic variation
- **Migration** affects genetic structure by mating between members of different populations
- **Genetic drift** is change in gene frequencies based on random events in finite populations
- **Natural selection** occurs in response to environmental stressors

Mutation

- Mutation is a rare event, so it has small quantitative effects but potentially large qualitative effects
  - Estimate of the average human genome mutation rate is ~2.5 x 10^{-8} mutations per nucleotide per generation (25 mutations per billion nucleotides per generation)
- However, even small changes to the sequence of nucleotides in the DNA can have significant repercussions in terms of protein structure and function

Sickle Cell Anemia

- Sickle cell anemia is a severe genetic condition caused by a **point** mutation: the change in one nucleotide within the sequence of 438 bases coding for the hemoglobin beta chain
- The transversion mutation in the 17th nucleotide from a Thymine base to an Adenine base causes a shift in the 6th amino acid from glutamic acid to valine

Sickle Cell Anemia

- The change of one amino acid results in hemoglobin that has a tendency to clump together and destroy Red Blood Cells

- This produces a life-threatening disease that has only come under good control by modern medicine in the last several decades

Point Mutation
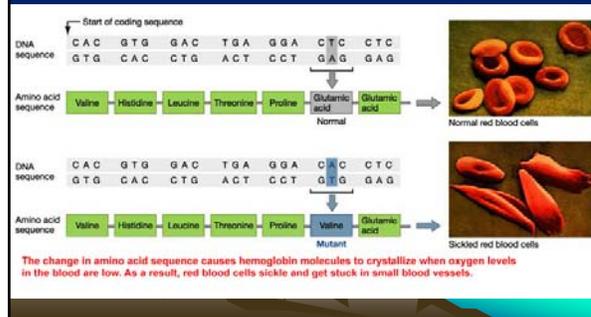
Migration or gene flow

- Gene Flow, the intermarriage or mixing between populations, has the effect of altering allele and genotype frequencies so that the two (or more) populations involved come to resemble each other in terms of genetic frequencies
  - Also known as admixture

Gene Flow

Population Boundaries

Mating between members of different populations

Individual
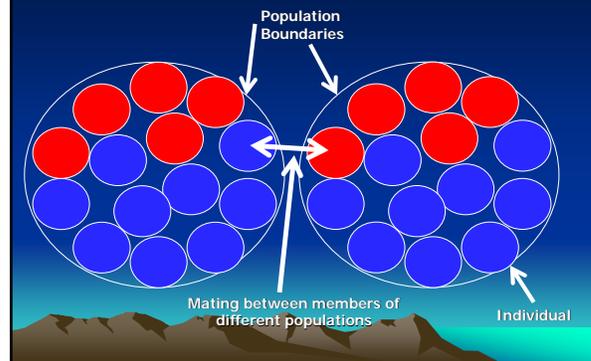
Admixture in African Americans

- Admixture has resulted in significant genetic changes in African Americans
- The Duffy blood group allele, Fy^a, was absent in enslaved Africans brought to the US
  - In West Africa the allele is close to 0%
  - The allele frequency is about 40% in US Whites
- Based on this, the % of white admixture in the African American population has been estimated:
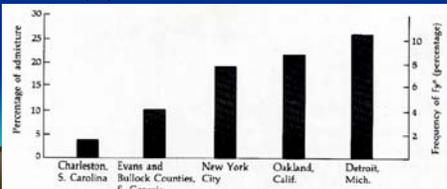
Selection

- Selection causes changes in allele and genotype frequencies from one generation to the next due to *differential net reproductive success* of individuals with different genotypes
  - If individuals with genotype AA consistently have twice as many offspring as individuals with AB and BB genotypes, the frequency of the A allele will increase and eventually, everyone will have the AA genotype

Selection

- There are two elements contributing to the differential reproductive success of individuals with differing genotypes
  - Viability or survival: individuals must survive to maturity in order to be able to reproduce
  - Fertility: individuals must produce offspring in order to pass on their genes
- The genotype producing the most offspring on average is the most fit

Natural Selection

- Natural selection is a two-step process
  - Production of variation (mutation)
  - Differential reproduction of favorable variants (selection)
    - Fitness is a measure of relative reproductive success by genotype
    - Ranges between 0 (die with no offspring) and 1 (die with maximum number of surviving offspring)
    - Selection (s) = 1 minus Fitness; represents selection against the genotype
      - Maximum selection against = 1 when fitness = 0

Selection against a recessive

Allele Frequency vs Generations of Selection

- s = 0.2
- s = 0.5
- s = 1
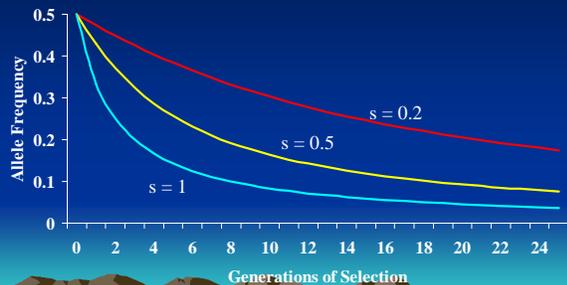
Random Effect in Finite Populations

- Genetic Drift
  - A finite population size is subject to chance or random influences on gene frequencies from one generation to the next
    - Smaller the population, the larger the effect
  - Founder effect is a small sample of a larger population that becomes a founder of a new population


Genetic Drift and Human Evolution

- Until recently, most humans lived in small, genetically isolated hunter-gather groups of perhaps 20-30 people
- Gene frequencies in groups of this size are highly susceptible to the effects of genetic drift
- As a result, much of the genetic variation in modern human populations may be a result of random processes such as genetic drift and natural selection

---

Bottleneck Effect

- This is a genetic sampling error that occurs when a catastrophic event reduces the size of a population to a few individuals
- The small number of survivors will almost certainly differ genetically from the ancestral population (e.g. rare alleles will be lost)
- The survivors will give rise to a new population that differs genetically from the one that existed before the catastrophe

After Bottleneck

Time

← Bottleneck

Before Bottleneck
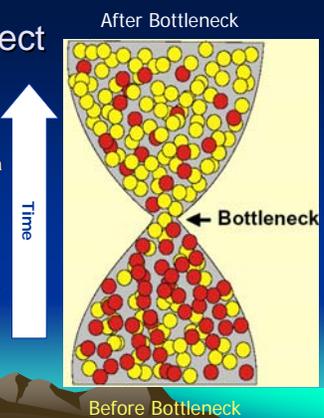
Founder Effect

- The genetic difference between the gene pool of a population and that of a newly isolated subpopulation from the same group
  - When a small number of individuals form a new population they have a random sample of the genes of the larger population they are derived from
    - Allele presence and frequencies may vary significantly from the parent population
- As the size of the founding population increases, the magnitude of the founder effect decreases
  - A recent estimate of the founding population of Native Americans was ~80 individuals!

Conclusions

- Genes are distributed in populations
- Populations differ from one another in the frequencies of different alleles
- The ultimate source of new genetic variation is mutation
- Gene flow makes population resemble one another
- Genetic drift and natural selection reduce variation in specific ways
  - Genetic drift is not related to environment, natural selection is