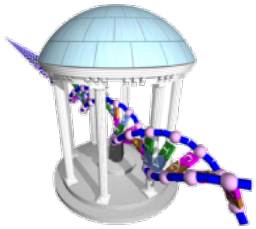
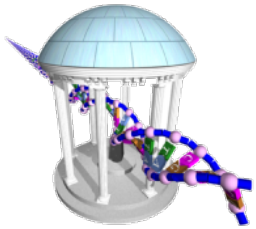


Phasing and Genomic Sequence

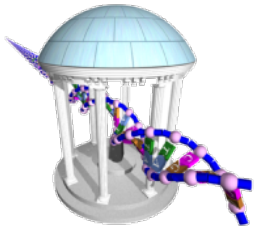
- Recessive and dominant traits
- Alleles live on linear sequences
- Inheritance of sequence
  - Recombination
  - Independent Segregation
- Trio Phasing



Traits

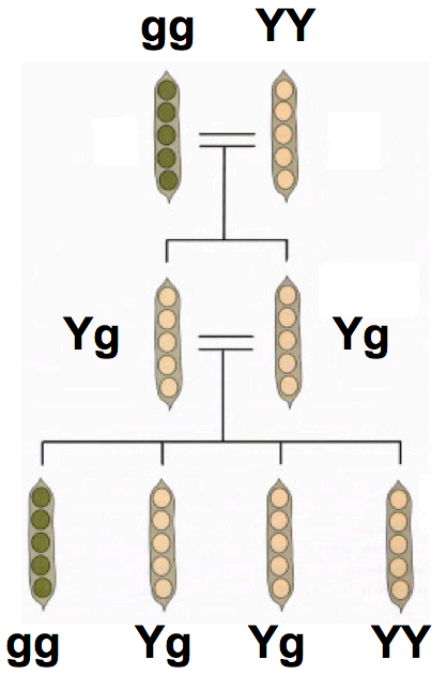
- Some phenotypes are directly correlated with a single allele, and thus simplify reasoning about them
  - What does the allele effect?
  - How common is an allele in nature?
  - What happens to the trait when specific crosses are made?
- Traits that are controlled by a single allele are call Mendelian Traits, whereas traits that are impacted by many alleles are called “Complex”
- In a diploid organisms the combination of alleles can manifest in various ways with regard to allelic state

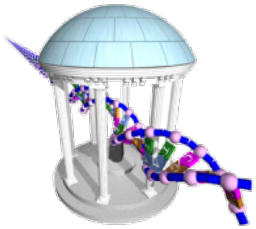
Gregor Mendel
1822-1884



Recessive Traits

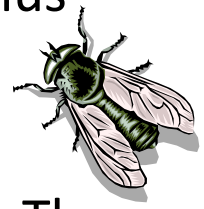
- Mendel discovered many important genetic properties using these simple traits
  - Traits are inherited by separately transferring the components allelic state to offspring (1st Law)
  - This mixing of allelic state is roughly *independent* (2nd Law)
  - Some allele combinations mask traits (makes them hidden, latent, or *recessive*), only to reappear in later generations
- However, violations of his 2nd law led to another discovery



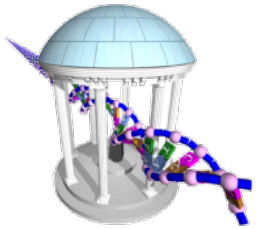


Early 20th Century Genetics

- In 1908, Thomas Hunt Morgan and Alfred H. Sturtevant showed that *genes* were organized in a linear sequence on chromosomes. Experimenting with Drosophila (fruit flies) they found sex chromosomes, sex-linked traits, and *crossing-over*. They were able to associate mutations to specific chromosomal regions, thus mapping gene locations.



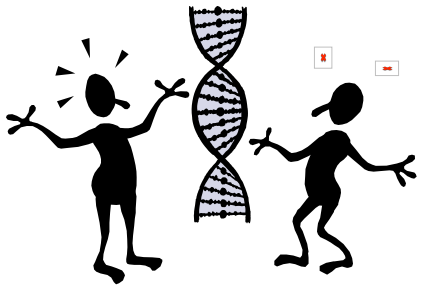
- By the 1930's biochemists knew that the nucleic acid present in chromosomes was DeoxyriboNucleic Acid, DNA. They also knew that chromosomes contained proteins in addition to DNA. DNA appeared to be long repetitive chains, and therefore, it seemed unlikely to carry information. Proteins, however, looked more interesting and were generally assumed to contain genetic materials. DNA was considered as just some sort of glue.

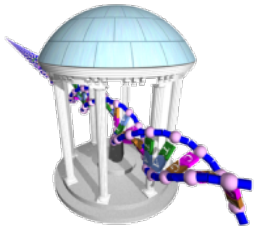


DNA's Central Role

- In 1944, Oswald Avery showed that DNA, not proteins, carries hereditary information.
- In the late 1940's and early 50's Linus Pauling and associates develop modeling methods for simultaneously determining structure and chemical make-up of proteins and other large molecules.
- In 1952, James Watson and Francis Crick, are able to determine the structure and chemical makeup of DNA, using X-ray crystallography data collected by Rosalind Franklin and Maurice Wilkins.

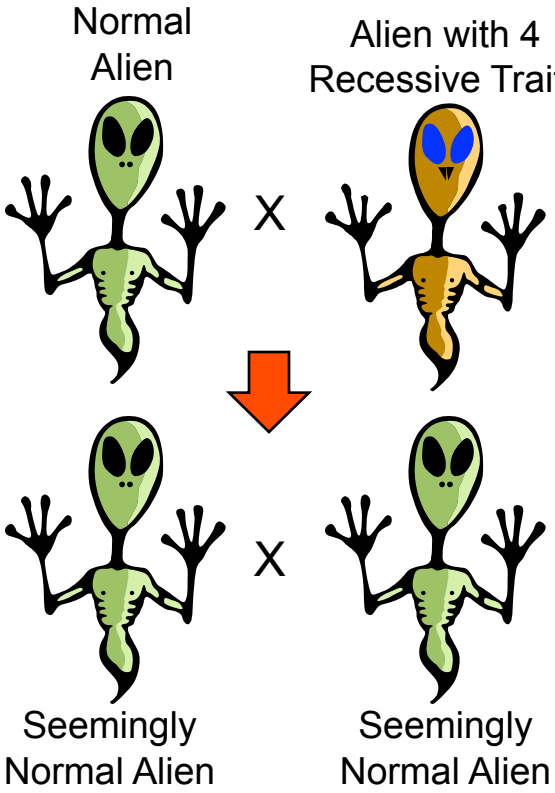
Beginning of Molecular Biology!



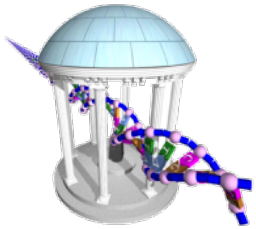


Inferring Genetic Maps preDNA

- Even without knowing the mechanisms of how heredity information is represented, clever scientists (Morgan) were able to "map" genes…

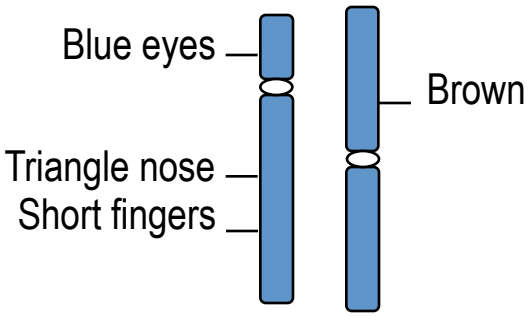


Normal	201	Short-fingered	9
Brown	64	Brown, blue-eyed, & triangle nosed	6
Blue-eyed	58	Triangle-nosed	5
Short fingered & triangle-nosed	54	Short-fingered & Brown	5
Short fingered, triangle-nosed, & brown	21	Brown, short fingered, triangle-nosed, & blue-eyed	4
Short fingered, triangle-nosed, & blue-eyed	20	Short-fingered & blue-eyed	4
Brown & blue-eyed	19	Triangle-nosed & brown	1
Blue-eyed & triangle-nosed	12	Brown, short fingered, & blue-eyed	1

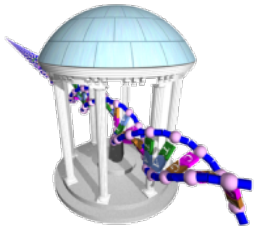


Steps to Infer a Genetic Map

- Verify Mendelian ratios of recessive traits
  - Brown (64 + 21 + 19 + 6 + 5 + 4 + 1 + 1) /484 = 0.250
  - Blue-eyes (58 + 20 + 19 + 12 + 6 + 4 + 4 + 1) / 484 = 0.256
  - Triangle-nose (54 + 21 + 20 + 12 + 6 + 5 + 4 + 1) / 484 = 0.254
  - Short-finger (54 + 21 + 20 + 9 + 5 + 4 + 4 + 1) / 484 = 0.244
- Test for pairwise linkages
  (we'd expect ¼ × ¼ × 484 ≈ 30 if independent)
  - Short-finger & triangle-nose 54 + 21 + 20 + 4 = 99
  - Triangle nose & brown 21 + 6 + 4 + 1 = 32
  - Short-finger & brown 21 + 5 + 4 + 1 = 31
  - Blue-eyes & triangle-nose 20 + 12 + 6 + 4 = 42
  - Short-finger & blue-eyes 20 + 4 + 4 + 1 = 29
  - Brown & blue-eyes 19 + 6 + 4 + 1 = 30
- Indicates
  - Short-fingers & triangle-nose are closely linked
  - Blue-eyes & triangle-nose are probably linked
  - Short-finger & blue-eyes appear independent, thus the triangle nose gene should lie between them
  - Brown gene is likely to be on another chromosome.



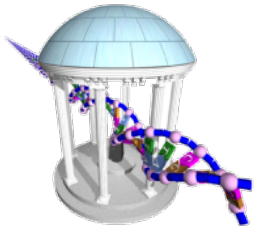
Morgan came up with even more clever techniques that were able to precisely locate the relative positions of genes on chromosomes. Even today chromosomal gene positions are measured in units of centiMorgans



Inheritance of Genetic Sequence

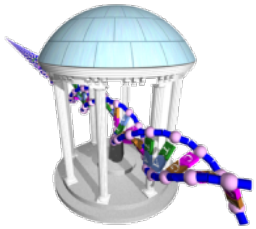
- Alleles are laid out on sequences, which are largely inherited intact
- Each parent passes one of its "sequences" to its children (Mendel's 1st law). At least for homologous chromosomes.
- How can this help us with phasing?

  Thus far we have considered each allele's state independently in our (Mom, Pop, Kid) trios. We can use an assumption of linkage (like Morgan) to infer what combinations of alleles co-occur in our pedigree, and use these short sequences to resolve Het|Het|Het cases.



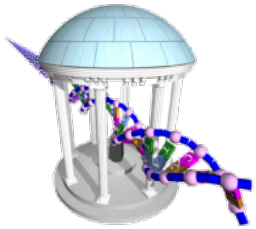
Halpotype Sequence

- The combination of alleles at adjacent locations that are inherited together are called a *haplotype*
- Diploid organisms with identical haplotypes on their homologous chromosomes will be homozygous at every allele, and are called inbred
- Haplotype diversity is necessary for
  - Phasing (why?)
  - Mapping traits
- However, in special cases (non-autosomes) inferring the haplotype sequence is simpler.



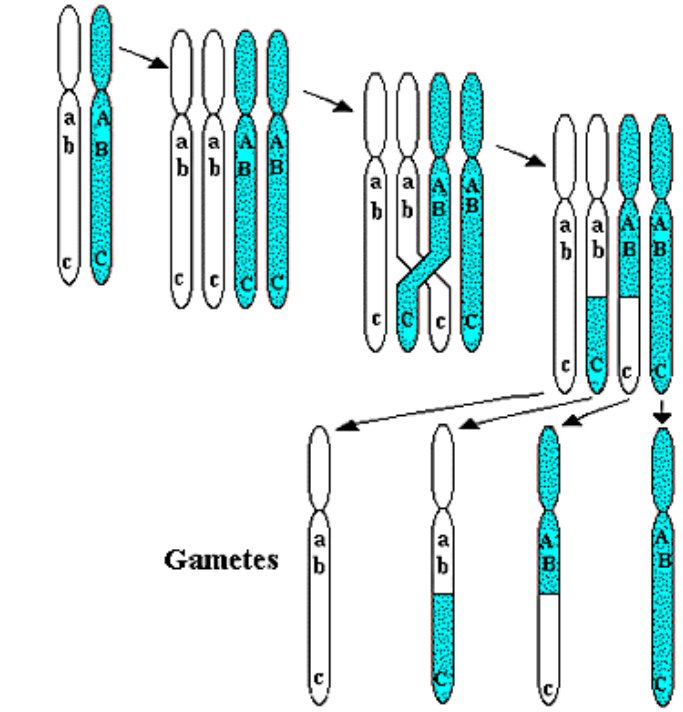
Simple Haplotypes

- Mitochondrial DNA:
  - Organelles in every cell
  - Single hapliod sequence inherited maternally
  - Short, less than 17 Kbases, ~37 genes
  - Present in both sexes
- Y DNA
  - Largely hapliod sequence inherited paternally, there is a small region on each end called the PAR that is an exception
  - Present only in males
  - Around 50 Mbases, relatively few genes
- X DNA
  - In females, one haplotype is inherited parternally largely intact aside from the PAR region (only need father-daughter pair to phase X)

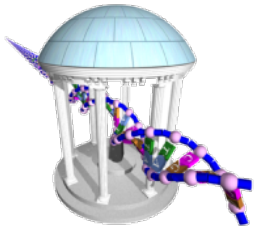


Crossover Recombination

- In the formation of gametes (sperm and ovum) homologous DNA strands are combined in a process called crossover
- This effectively combines the prefix of one sequence with the suffix of another

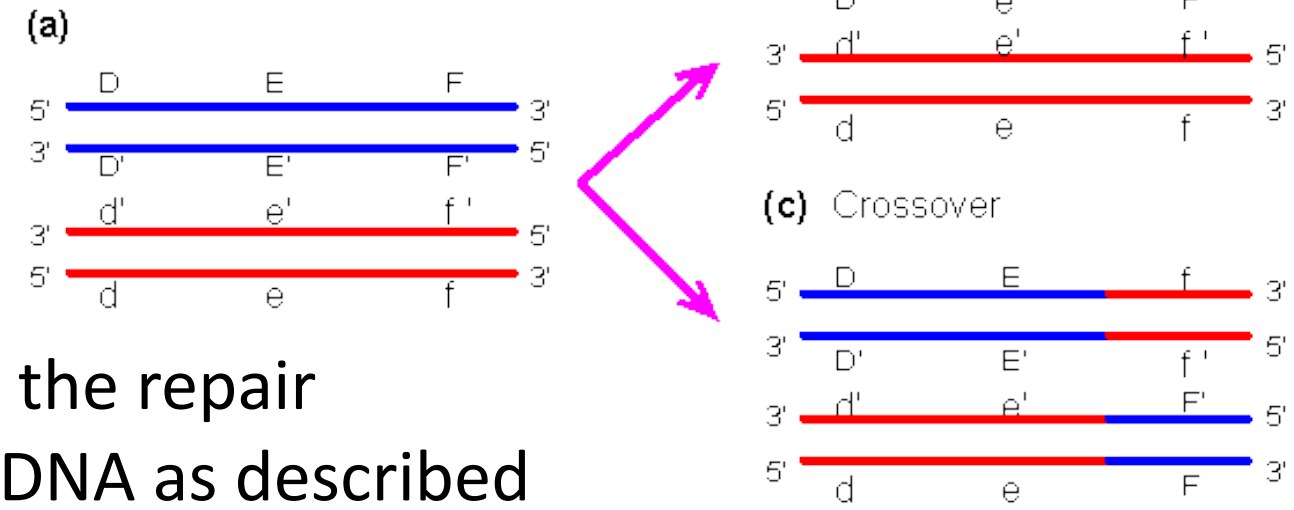


Crossing-over and recombination during meiosis

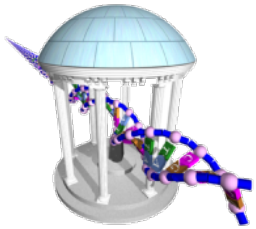


Gene Conversion Recombination

- The DNA sequence is transferred from one copy (which remains unchanged) to another, whose sequence is altered.

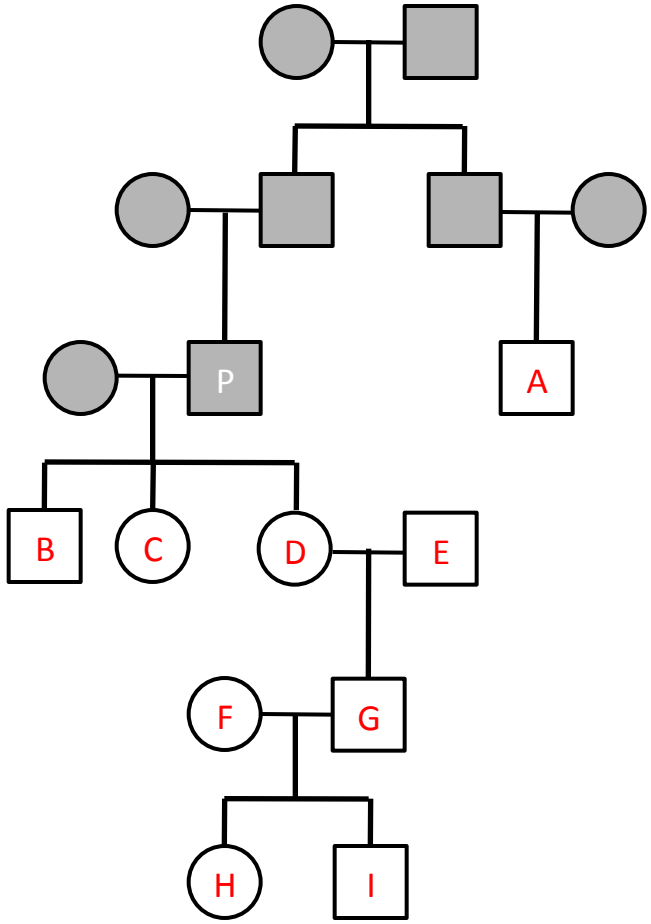


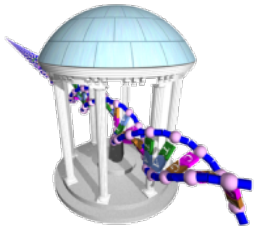
- Results from the repair of damaged DNA as described by the Double Strand Break Repair Model.



Concrete Example

- Return to our pedigree
- Examine the haplotype diversity of the 6 male samples (A,B,E,G,I,Q)
- Choose any of the unphased het|het|het loci that you found during Trio phasing and examine the population of genotypes for a common haplotype that might resolve it.





Next Time

- There is no class meeting next Monday due to the holiday
- I need to attend a PhD defense held next Wednesday, and I would like for all of you to attend it rather than class if possible.

John Didion

Ph.D. Thesis Defense
Noon, Pagano Conference Room (Lineberger)