03-621 Week 6
Advanced Quantative Genetics

Aidan Jan

February 20, 2026

Mapping Traits onto the Genome Sequence

Knowing the human genome sequence is not enough to identify specific genes nor their biological functions. A genetic map is what makes genome sequences **interpretable**.

- Knonowing the location of a gene allows to identify the DNA sequence corresponding to a gene, and the DNA sequence variations that correspond with different alleles.

Basic Techniques for Mapping

For linkage and recombination:

- For linkage disequilibrium: use mapping by association. This works for complex traits and single-gene traits
- Map by recombination frequency for single gene traits
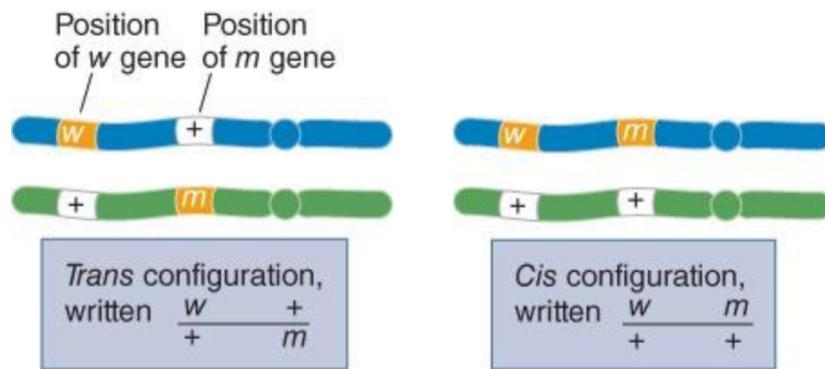
For chromosome abnormalities, use cytological mapping, which reveals single-gene traits.

Nomenclature

- $+/+$, a^+/a^+, A^+/A^+: wild type homozygote
- a/a: mutant homozygote (a recessive to wild type)
- $a/+$, a/A^+: heterozygote (exhibits WT phenotype)
- A/A: mutant homozygote (A dominant to wild type)
- $A/+$, A/a^+: heterozygote (exhibits A phenotype)
- $-/-$: homozygote for a null allele
- $-/+$: heterozygote for a null allele

A null allele is one that has lost function completely. Usually recessive to WT.

Nomenclature for allele arrangements

Recombination

Recombination is the production of new combinations of alleles during meiosis

- For genes on the **same chromosome**, recombination is a consequence of crossing over during Meiosis 1

- Crossing over involves the physical exchange of DNA segments between homologs, at chiasmata. (A chaisma is the physical location where the chromosomes cross.)

- Note: Chiasmata are generally required for proper segregation of homologs during Meiosis I.

The observed frequency of recombinant chromosomes is a measure of the distance between the two genes:

$$\text{Genetic distance} \ = 100 \times \frac{\text{Recombinants}}{(\text{Recombinants} \ + \ \text{Parentals})}$$

Example

Consider this dihybrid test cross involving recessive color and wing size traits in Drosophila (house flies)

$$\text{Wild type (Gray with normal wings)} \times \text{Mutant (Black with vestigal wings)}$$

Offspring:

- Wild type: 965

- Black-vestigial: 944

- Gray-vestigial: 206

- Black-normal: 185

Are b and vg linked? If there is independent assortment, what phenotypic ratio do we expect?

- If independent assortment, we expect $1 : 1 : 1 : 1$.

Chi-squared (χ^2) Test for Independent Assortment

1. Define the null hypothesis: b and vg assort independently

2. Collect data. (Shown above.)

3. Calculate expected progeny ratios and numbers. (Expected is each one has a count of 575, calculated by averaging the numbers.)

4. Calculate Chi-square metrics: $\chi^2 = \sum[(\text{observed} - \text{expected})^2/\text{Expected}]$

$$\chi^2 = \frac{(965 + 944 - 1150)^2}{1150} + \frac{(206 + 185 - 1150)^2}{1150} = 501 + 501 = 1002$$

5. Determine degrees of freedom: $df = n - 1$. (2 - 1 = 1)

6. Calculate P: (Excel: CHISQ.DIST.RT. Alternatively, refer to a table).

   - We get $P(\chi^2 = 1002|df = 1) = 6.6 \times 10^{-220}$

For our example, the χ value is tiny compared to the null hypothesis $\alpha = 0.05$. Therefore, we reject the null hypothesis. The data are not compatible with independent assortment of b and vg. Instead, they support linkage on the same chromosome.

Calculating Genetic Distance
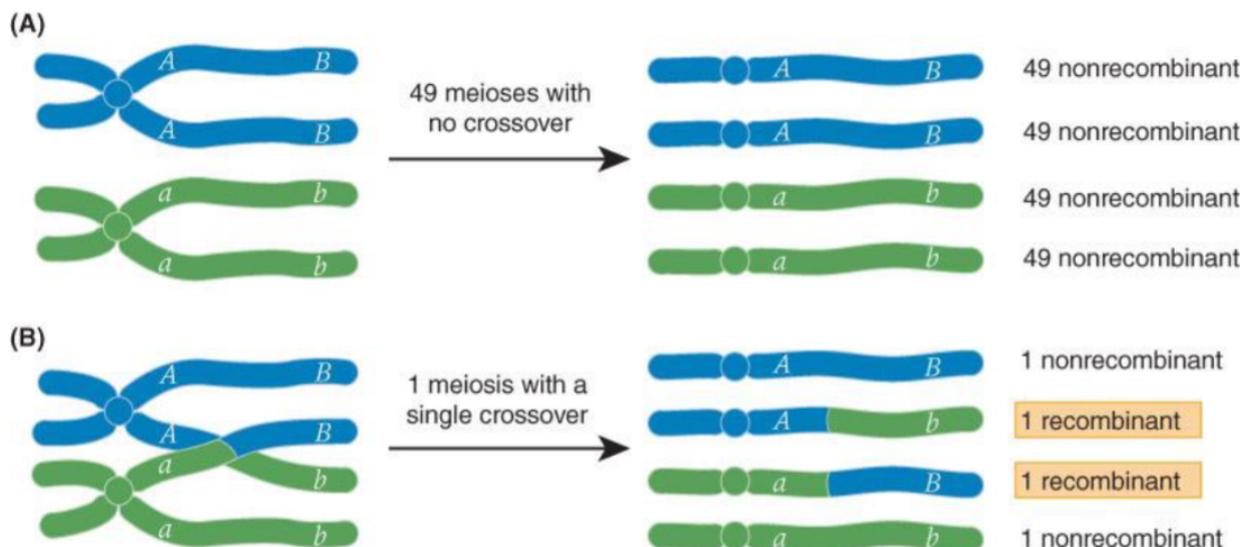
Given b and vg are linked, what is the genetic distance between them?

$$100 \times \frac{206 + 185}{965 + 944 + 206 + 185} = 0.17$$

The units are called **centiMorgans** (cM) after T.H. Morgan. 1cM = 1% recombinants. (This does not imply 1% crossing over!)

- Genetic distance are essentially a map of the physical distance two genes are separated on a chromosome.

- Genetic distances based on recombination frequencies allows us to map the relative positions of genes on chromosomes.

- Over short intervals ($<\sim 20$ cM): map distances are often additive. Over long intervals, % recombinants underestimate the map distance.

- In humans, 1 cM =$\sim$1 million base pairs.

- Varies locally, but on average, 1 million base pairs.

- Differs between sexes ($\sim 2\times$) higher in females

Biological Meaning of Map Distance

$$r = \frac{1+1}{49 + 49 + 49 + 49 + 1 + 1 + 1 + 1} = \frac{2}{200}$$

Thus, a 1 cM interval is a chromosomal segment that undergoes 1 crossover event per 50 meioses.

Because even numbers of multiple crossovers restore the parental marker configurations, the maximum expected recombination frequency is 50%.
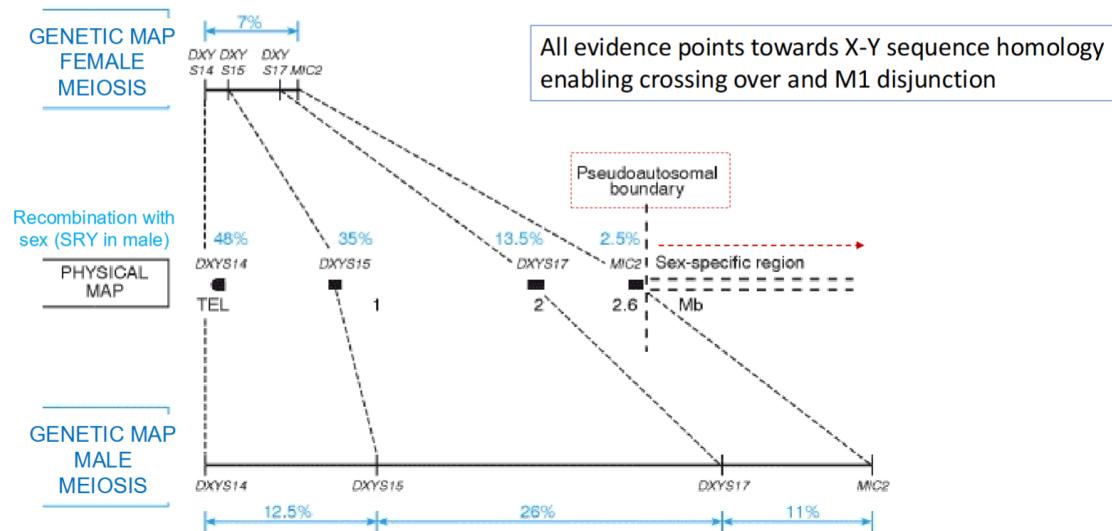
Recap: Genetic Map Distance

- **Scale**: frequency of cross-over (i.e., recombinant) gametes, which is a positive function of the physical distance between two loci

    - Unit: centi-Morgan (cM)
    - 1 cM = 1 crossover per 50 meioses
    - This corresponds to a recombinant frequency $\theta = 1\%$

- With the genome sequence in hand, the physical distance can be measured in nucleotides, so the genetic distance is an indication of the number of nucleotides between two loci.

    - The correlation allows us to identify the gene sequences for specific traits of interest.

X-Y Crossover

Pseudoautosomal region 1 (PAR1) is a section on the human X and Y chromosomes where recombination is guaranteed to occur in male gamete meiosis.
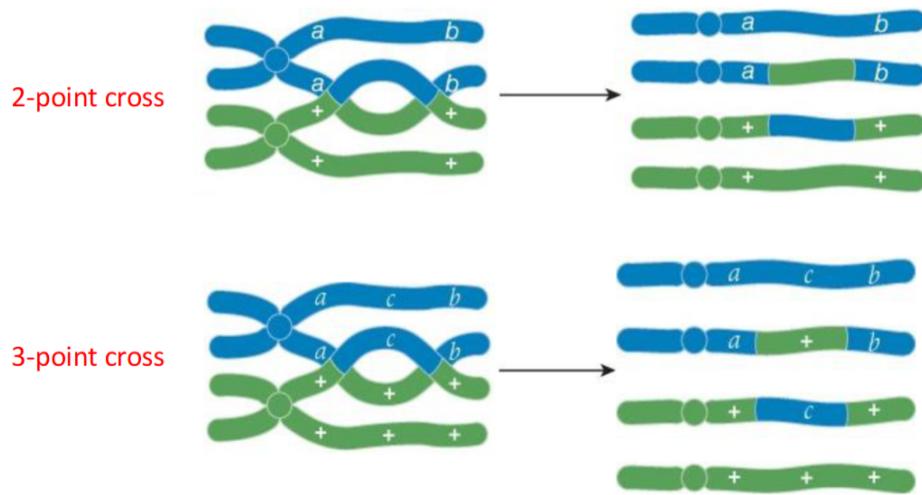
- It is a gene that is shared on both chromosomes, albeit in different locations

Problems over large distances

We cannot detect double cross overs because it does not result in recombination between A and B genes

- We use **three-point crosses** to estimate linkage over **long distances** and determine marker order

2-point cross

3-point cross

In real life, we often **don't know** the order of the markers; that is what we want to find out! One way to do this is by calculating the map distances, and then deciding which of the possible map orders fits.

- Another way is by realizing that the double X-over progeny classes are much less frequent than any single X-over progeny class.

Shortcut to determining gene order in a 3-point cross

The gene in the middle is always the one that switches **both** partners in the double crossover.

Do double crossover events occur as often as expected?

- Generally, no

- "Interference": reduction in numbers of double crossover events observed as compared to expected.

- The expectation is calculated from the frequencies of single crossover events.

$$\text{Coefficient of Coicidence} = \frac{\text{Observed \# of double recombinant chromosomes}}{\text{Expected \# of double recombinant chromosomes}}$$

$$\text{Interference} = 1 - \text{ Coefficient of Coincidence}$$

Population Genetics

Studies the frequencies of alleles and genotypes at the level of populations (natural or domesticated), where:

- We may not be able to control crosses (genotypes of parents)

- Genealogical relationships among individuals may be unknown

Population: The potentially interbreeding individuals in a given geographical area (they contribute to a common gene pool via their gametes)

Population Genetics Questions

- How rapidly do allele frequencies change under selection?

- Why do harmful alleles persist in a population?

- Why are some single-gene disorders rare and others comparatively common?

- Why are some single-gene disorders common in some populations but rare in others?

Allele frequencies in populations are determined by:

- Mutation

- Selection

- Sampling Variance

The Hardy-Weinberg Principle

Recall: **If mating is random** and there are **two alleles** at a given locus with frequencies p and q, the distribution of genotype frequencies in the population is given by the binomial expression:

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

This all works **if** certain assumptions are true . . .

The **Gene Pool** refers to the total of all alleles at all genes in the population.

- Let the probability of drawing a A gamete (sperm or egg) be p

- Let the probability of drawing a a gamete (sperm or egg) be q

Hardy-Weinberg Assumptions

Hardy Weinberg gives the mathematical relationship between allele and genotype frequencies in an *idealized* population

1. Mating is random

2. All genotypes are equal in <u>viability</u> and <u>fertility</u> (no selection)

3. Mutation does not occur

4. Migration into the population does not occur (no gene flow)

5. The population is sufficiently large that sampling variation during reproduction does not cause random changes in allele frequency (no drift)

The Hardy-Weinberg Principle ties together the genotype frequencies

- For what allele frequencies is the frequency of heterozygotes maximal? (0.5)

- For very rare alleles, the proportion found in **heterozygotes** far exceeds the proportion in homozygotes

Sex Linkage on Recessive Frequency

- The **allele frequencies** in X-bearing male gametes are the same as in female gametes (p and q).

- The **allele frequencies** in male and female progeny are also the same (p and q)

- But the frequency of the **recessive phenotype** is **higher** in male progeny than female progeny.

For example, the *G6PD* gene in humans is X-linked. Recessive alleles cause haemolytic anemia that is exacerbated by diet (e.g., fava beans)

- In a certain population, 0.0025 of newborn males are affected.

- What is the frequency of disease alleles in that population?

    - Since males must inherit a Y chromosome from their father, the disease must come from the mother. Therefore, the proportions must be $p = 1 - 0.0025 = 0.9975$, and $q = 0.0025$.

- What proportion of newborn femals are carriers?

    - Since we have the values of p and q, heterozygotes are given by $2pq$, so therefore 0.00249 of newborn females are carriers.

Extending the Hardy-Weinberg Principle

We can use Hardy-Weniberg for more than two alleles. If there are n alleles, the genotype frequencies are given by the polynomial:
$$1 = (p_1 + p_2 + p_3 + \cdots + p_n)^2$$
For example, if there are three alleles, we have:

$$1 = (p_1 + p_2 + p_3)^2 = p_1^2 + p_2^2 + p_3^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3$$

Shortcuts for a Rare Recessive Trait

If the trait is less than 1 in 1000, then $p \approx 1$. Therefore,

- Frequency of carriers in the population: $2pq \approx 2(1)q = 2q$

- Probability that a random *unaffected* individual is a carrier: $2pq/(1 - q^2) \approx 2(1)q/1 = 2q$

- Probability that a random *unaffected* individual is homozygous dominant: $p^2/(1 - q^2) \approx p^2/1 = p^2$