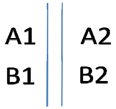
Introduction to Linkage

Law of Independent Assortment

- Mendel's Second Law (Law of Independent Assortment) :
  - The segregation of the genes for one trait is independent of the segregation of genes for another trait, i.e., when genes segregate, they do so independently
- This law essentially states that during gamete formation, the segregation of one gene is independent of the other gene
- This "law" is frequently violated and is only true for loci/genes that are unlinked.

Recombination

- When a gamete is passed down, the chromosome inherited by an offspring from a parent is actually a mosaic of the parent's two chromosomes.
- Suppose we have two loci on the same chromosome, locus 1 and locus 2, where locus 1 has alleles A1 and A2, and locus 2 has alleles B1 and B2.
- In the example below, **phase** is known and is (A1,B1) and (A2,B2).
- If the genes are closely linked, a gamete is much more likely to contain (A1,B1) or (A2,B2), which are "non-recombinants."
- If there is recombination, a gamete will contain (A1, B2) or (A2,B1), but this is less likely if the loci are linked.



Recombination Fraction

- Two loci that are unlinked follow Mendel's Second Law, and all possible gametes for a parent are produced with equal frequency.
- When loci are physically located close to one another on a chromosome, there is a deviation from this relationship. This deviation is summarized by the recombination fraction.
- The recombination fraction is often denoted by θ where $0 \leq \theta \leq \frac{1}{2}$
- P(recombinant gamete)$= \theta$
- If $\theta < \frac{1}{2}$, the loci are said to be linked or in genetic linkage
- When loci are completely linked, $\theta = 0$
- Two loci are said to be unlinked if $\theta = \frac{1}{2}$.
- Note that if two loci are on different chromosome, then $\theta = \frac{1}{2}$.

Linkage in a simple genetic cross

- In the early 1900's, Bateson and Punnet conducted genetic studies using sweet peas. They studied two characters:
  - Petal color which has two alleles: P (purple) and p (red), where P is dominant.
  - Pollen grain shape has two alleles: L (elongated) and l (disc-shaped), where L is dominant

$$\text{PPLL} \times \text{ppll}$$
$$\downarrow$$
$$\text{PpLl} \qquad\qquad \text{F1}$$

- Plants in the F1 generation were intercrossed: PpLl X PpLl.
- According to Mendel's Second Law, during gamete formation, the segregation of one gene pair is independent of another gene pair.
- If this genetic model is correct, what segregation ratios of the phenotypes in the F2 generation would we expect?

Sweet Peas Linkage Example

F2	PL	Pl	pL	pl
PL	Purple/Long	Purple/Long	Purple/Long	Purple/Long
Pl	Purple/Long	Purple/Disc	Purple/Long	Purple/Disc
pL	Purple/Long	Purple/Long	Red/Long	Red/Long
pl	Purple/Long	Purple/Disc	Red/Long	Red/Disc

Sweet Peas Linkage Example

- The expected relative frequencies in the F2 generation if the genes segregated independently are

	Elongated	Disc-Shaped
Purple	9	3
Red	3	1

- The observed frequencies in 381 plants in the F2 generation where

	Elongated	Disc-Shaped
Purple	284	21
Red	21	55

- The observed data clearly do not fit what is expected under the model.
- The explanation: the petal color gene and the gene for pollen grain shape are linked.
- Let θ be the recombination fraction between the two genes. What is the probability of each possible plant type?

Sweet Peas Linkage Example

		$\frac{1}{2}(1-\theta)$	$\frac{1}{2}\theta$	$\frac{1}{2}\theta$	$\frac{1}{2}(1-\theta)$
		PL	Pl	pL	pl
$\frac{1}{2}(1-\theta)$	PL	Purple/Long	Purple/Long	Purple/Long	Purple/Long
$\frac{1}{2}\theta$	Pl	Purple/Long	Purple/Disc	Purple/Long	Purple/Disc
$\frac{1}{2}\theta$	pL	Purple/Long	Purple/Long	Red/Long	Red/Long
$\frac{1}{2}(1-\theta)$	pl	Purple/Long	Purple/Disc	Red/Long	Red/Disc

- P(red, disc-shaped)= $p_{R/D} = \frac{1}{4}(1-\theta)^2$
- P(red,elongated)= $p_{R/L} =$
  $\left(\frac{1}{2}\theta\right)\left(\frac{1}{2}\theta\right) + \left(\frac{1}{2}\theta\right)\left(\frac{1}{2}(1-\theta)\right) + \left(\frac{1}{2}(1-\theta)\right)\left(\frac{1}{2}(\theta)\right) = \frac{1}{4}\theta(2-\theta)$
- P(purple, disc-shaped)= $p_{P/D} = \frac{1}{4}\theta(2-\theta)$
- P(purple, elongated)= $p_{P/L} = \frac{1}{2} + \frac{1}{4}(1-\theta)^2$
- We can form a likelihood for the data that is a function of the recombination fraction θ. We can find the value of θ that maximizes this likelihood.

Sweet Peas Linkage Example Likelihood: Multinomial Distribution

- The likelihood function will have the following multinomial distribution:

$$L\left(n_{P/L}, n_{P/D}, n_{R/L}, n_{R/D} \middle| p_{P/L}, p_{P/D}, p_{R/L}, p_{R/D}\right)$$

$$= \binom{N}{n_{P/L}, n_{P/D}, n_{R/L}, n_{R/D}} \times$$

$$(p_{P/L})^{n_{P/L}} (p_{P/D})^{n_{P/D}} (p_{R/L})^{n_{R/L}} (p_{R/D})^{n_{R/D}}$$

Sweet Peas Linkage Example Likelihood: Multinomial Distribution

- Obtain the log-likelihood function to

$$l\left(n_{P/L}, n_{P/D}, n_{R/L}, n_{R/D} \middle| p_{P/L}, p_{P/D}, p_{R/L}, p_{R/D}\right)$$

$$= C + n_{P/L} ln(p_{P/L}) + n_{P/D} ln(p_{P/D}) + n_{R/L} ln(p_{R/L}) + n_{R/D} ln(p_{R/D})$$

where

$$C = ln\left[\binom{N}{n_{P/L}, n_{P/D}, n_{R/L}, n_{R/D}}\right]$$

- Remember that $p_{P/L}, p_{P/D}, p_{R/L}, p_{R/D}$ are all functions of θ.
- Can obtain the maximum likelihood estimate of θ by taking the derivatives of the log-likelihood function with respect to θ and solving for 0.