Linkage and crossing over

Deependra Dhakal

Academic year 2019-2020

Gokuleshwor Agriculture and Animal Science College
Tribhuwan University
ddhakal.rookie@gmail.com
https://rookie.rbind.io

Linkage and crossing over

Locating genes in chromosomes

- Reasons:
    - Gene position is crucial information needed to *build complex genotypes* required for experimental purposes or for commercial applications.
    - Knowing the position occupied by a gene provides a way of discovering its *structure and function*.
    - Chromosomal arrangements of related species might differ. By comparing such differences, geneticists can deduce the *evolutionary genetic mechanisms* through which these genomes diverged.
- The arrangement of genes on chromosomes is represented digaramatically as a unidimensional chromosome map, showing gene positions, known as **loci**.
- Two basic types of chromosome maps are available:
    - Recombination map
    - Physical map

Recombination

- The independent assortment of genes at meiosis is one of the main ways by which an organism produces new combinations of alleles. The production of new allele combinations is formally called recombination.
- Meiotic recombination is any meiotic process that generates a haploid product with new combinations of the alleles carried by the haploid genotypes that united to form the meiocyte.
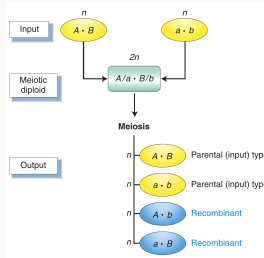

Figure 1: Recombinants (blue) are those products of meiosis with allele combinations different from those of the haploid cells that formed the meiotic diploid (yellow). Note that genes A /a and B /b are shown separated by a dot because they may be on the same chromosome or on different chromosomes.

Linkage

- Recombination maps of chromosomes are usually assembled two or three genes at a time, with the use of a method called linkage analysis.
- When we say two genes are **linked**, it means that the loci of those genes are on the same chromosome.
- In early 1900s, William Bateson and R. C. Punnett were studying the inheritance of two genes in sweet peas.
- In a standard self of a dihybrid F_1, the F_2 did not show the 9:3:3:1 ratio predicted by the principle of independent assortment. Bateson and Punnett noted that certain combinations of alleles show up more than expected, as if they were physically attached in some way.
- Thomas H Morgan found similar deviation from mendel's second law while studying autosomal genes in *Drosophila*.
- Morgan proposed linkage as a hypothesis to explain the phenomena of apparent allele association.

Linkage: Morgan's data

- One of the genes affected eye color (pr, purple, and pr^+, red), and the other gene affected wing length (vg, vestigial, and vg^+, normal). (Vestigial wings are very small compared to wild type.) The wild-type alleles of both genes are dominant.
- Morgan performed a cross to obtain dihybrids and then followed with a testcross:

$$\text{P} \qquad\qquad pr/pr.vg/vg \times pr^+/pr^+./vg^+/vg^+$$

$$\downarrow$$

Gametes $\qquad\qquad pr.vg \ pr^+ vg^+$

F_1 dihybrid $\qquad\qquad pr^+/pr.vg^+/vg$

Testcross:

$$pr^+/pr.vg^+/vg \female \times pr/pr.vg/vg \male$$

F_1 dihybrid femaleTester male

- Because the tester parent contributes gametes carrying only recessive alleles, the phenotypes of the offspring directly reveal the alleles contributed by the gametes of the dihybrid parent.
- Hence, the analyst can concentrate on meiosis in one parent (the dihybrid) and essentially forget about meiosis in the other (the tester).
- In contrast, from an F_1 self, there are two sets of meioses to consider in the analysis of progeny: one in the male parent and the other in the female.
- Morgan's testcross results were as follows (listed as the gametic classes from the dihybrid):

Genotype	Progenies
$pr^+.vg^+$	1339
$pr.vg$	1195
$pr^+.vg$	151
$pr.vg^+$	154

- These numbers drastically deviate from the Mendelian prediction of a 1:1:1:1 ratio expected from independent assortment (approximately 710 in each of the four class).
- Another useful way of assessing the testcross results is by considering the percentage of recombinants in the progeny.
- There are two recombinations in the cross, $pr^+.vg$ and $pr.vg^+$, because they are clearly not the two input genotypes contributed to the F_1 dihybrid by the original homozygous parental flies (by their gametes, in fact).
- Two recombinant types are approximately equal in frequency (151 ~ 154). Their total is 305 (frequency $= \frac{305}{2839} \times 100 = 10.7\%$).

- This data can be explained by postulating that the genes were linked on the same chromosome, and so the parental allelic combinations are held together in the majority of progeny.
- In the dihybrid, the allelic conformation must have been as follows:

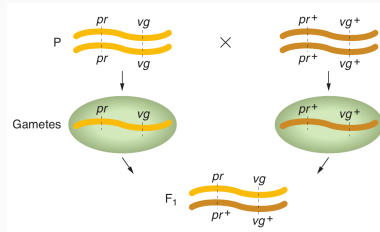$$\frac{pr^+ \qquad vg^+}{pr \qquad vg}$$

Figure 2: Inheritance of linked genes; These genes tend to be inherited as a package.

- Morgan made yet another cross with the use of the same alleles but in a different combination. In this cross, each parent is homozygous for the wild-type allele of one gene and the mutant allele of the other.
- Again F_1 females were testcrossed.
- Now,
    - Speak about progeny frequencies
    - Do the progeny frequencies match to Mendelian ratio ?
    - What are the recombinant classes ?
    - What is the likely conformation of the F_1 dihybrid ?
- Two forms of dihybrid testcross ratios are possible in genetics:
    1. Two equally frequent nonrecombinant classes totaling in excess of 50 percent
    2. Two equally frequent recombinant classes totaling less than 50 percent.

When two genes are close together on the same chromosome pair (that is, they are linked), they do not assort independently but produce a recombinant frequency of less than 50 percent.

Crossing over

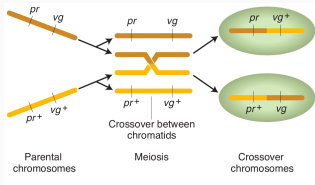## Crossing over produces new allelic combinations

Figure 3: Crossing over phenomena

- Crossover products are a result of break and join of two homologous chromosomes.

- At meiosis, when duplicated homologous chromosomes pair with each other – in genetic terms, when the two dyads unite as bivalent – a cross shaped structure called chiasma (pl., chiasmata) often forms between two nonsister chromatids.

- In 1931, Harriet Creighton and Barbara McClintock were studying two genes of corn that they knew were both located on chromosome 9. One affected seed color (C, colored; c, colorless), and the other affected endosperm composition (Wx, waxy; wx, starchy). The plant was a dihybrid in cis conformation.

Linkage symbolism

- Two conformations of linked genes in a dihybrid are possible.
    1. Cis : AB/ab or $++/ab$
    2. Trans: Ab/aB or $+b/a+$
- Alleles on the same homolog have no punctuation between them.
- A slash symbolically separates two homologs
- Alleles are always written in the same order on each homolog
- Genes known to be on different chromosomes are shown separated by a semicolon. For e.g., A/a; C/c

Two point crossing over

- As already evident from Morgan's data above, the recombinant frequency was significantly less than 50%, specifically 10.7%.
- Recombination frequencies greater than 50% is never observed.
- Imagine two specific genes positioned a certain fixed distance apart. Now imagine random crossing over along the paired homologs. In some meioses, nonsister chromatids cross over by chance in the chromosomal region between these genes; from these meioses, recombinants are produced. In other meiotic divisions, there are no crossovers between these genes; no recombinants result from these meioses.

- Sturtevant postulated a rough proportionality: the greater the distance between the linked genes, the greater the chance of crossovers in the region between the genes and, hence, the greater the proportion of recombinants that would be produced. Thus, by determining the frequency of recombinants, we can obtain a measure of the map distance between the genes.
- Sturtevant defined one genetic map unit (m.u.) as that distance between genes for which 1 product of meiosis in 100 is recombinant. For example, the recombinant frequency (RF) of 10.7 percent obtained by Morgan is defined as 10.7 m.u. A map unit is sometimes referred to as a centimorgan (cM) in honor of Thomas Hunt Morgan.

Additivity of genetic distance

- Does the method of genetic map unit assignment correspond to chromosome linearity ?
- Sturtevant predicted that, on a linear map, if 5 map units (5 m.u.) separate genes A and B, and 3 m.u. separate genes A and C, then the distance separating B and C should be either 8 or 2 m.u.

Three point crossing over

- Cross of a trihybrid (triple heterozygote) with a triply recessive tester gives rise to three-point testcross or a three-factor cross.
- The goal is to deduce whether the three genes are linked and, if they are, to deduce their order and the map distances between them.
- (Example from *Drosophila* mutant study) The mutant alleles are v (vermilion eyes), cv (crossveinless, or absence of a crossvein on the wing), and ct (cut, or snipped, wing edges). The analysis is carried out by performing the following crosses:

$$P \qquad v^+/v^+.cv/cv.ct/ct \times v/v.cv^+/cv^+.ct^+/ct^+$$
$$\downarrow$$
$$\text{Gametes} \qquad v^+.cv.t \quad v.cv^+.ct^+$$
$$\text{F}_1 \text{ trihybrid} \qquad v/v^+.cv/cv^+.ct/ct^+$$

- Trihybrid females are testcrossed with triple recessive males:

$$v^+/v^+.cv/cv.ct/ct♀ \quad ×v/v.cv^+/cv^+.ct^+/ct^+♂$$

$$\text{F}_1 \text{ trihybrid female} \quad \text{Tester male}$$

- From any trihybrid, only $2 \times 2 \times 2 = 8$ gamete genotypes are possible.
- They are the genotypes seen in the testcross progeny.
- Based on observation of progeny, following genotypes have been noted (Table **??**) at given frequencies.
- The columns alongside show which genotypes are recomninant (R) for the loci taken two at a time.
- Note: any combination other than orignal parental input genotypes are recombinant types.

18

| Gametes	Progenies	Recombinant for loci		
		v and cv	v and ct	cv and ct
$v.cv^+.ct^+$	580			
$v^+.cv.ct$	592			
$v.cv.ct^+$	45	R		R
$v^+.cv^+.ct$	40	R		R
$v.cv.ct$	89	R	R	
$v^+.cv^+.ct^+$	94	R	R	
$v.cv^+.ct$	3		R	R
$v^+.cv.ct^+$	5		R	R

- Starting with the v and cv loci, we have parentals for this pair of loci are $v^+.cv$ and $v.cv^+$, recombinants are $v.cv$ and $v^+.cv^+$.
- There are $45+40+89+94 = 268$ of these recombinants. Of a total of 1448 flies, this gives an RF of 18.5 %.
- For the v and ct loci, the recombinants are $v.ct$ and $v^+.ct^+$. There are $89+94+3+5 = 191$ of these recombinants among 1448 flies, and so the RF $= 13.2$ %.
- For ct and cv loci, the recombinants are $cv.ct^+$ and $cv^+.ct$. There are $45+40+3+5 = 93$ of these recombinants among all flies, and so the RF $= 6.4$ percent.

1. Are loci linked ?

- Because the RF values are all considerably less than 50 percent the loci are linked.

2. How are the loci ordered ?

- Because the *v* and *cv* loci have the largest RF value, they must be farthest apart; therefore, the *ct* locus must lie between them. A map can be drawn as follows:
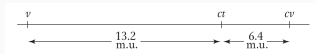
- First, we have deduced a gene order that is different from that used in our list of the progeny genotypes.
- The original listing was assumed; the order was simply not known before the data were analyzed
- We have established that *ct* is between *v* and *cv*.
- But could the map be inverted ?
- Linkage maps merely map the loci in relation to one another, with the use of standard map units, which we have done based on recombinants' frequency on the measure of map units.

Double crossover

- Point to note is that the two smaller map distances, 13.2 m.u. and 6.4 m.u., add up to 19.6 m.u., which is greater than 18.5 m.u., the distance calculated for *v* and *cv*.
- This is because of how we have treated rarest recombinants with respect to the recombination of *v* and *cv*.
- These individuals (recombinants of outer loci) are in fact double recombinants, arising from two crossovers.
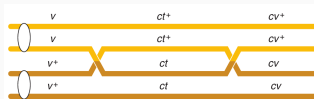
- When we calculated the RF value for v and cv, we did not count the $v.ct.cv^+$ and $v^+.ct^+.cv$ genotypes; after all, with regard to v and cv , they are parental combinations.
- In light of our map, however, we underestimated the distance between the v and cv loci.
- We should have counted the two rarest classes, and included them twice because each represents double recombinants.
- Hence we can correct the value by adding the numbers $45+40+89+94+3+3+5+5 = 284$. Of the total of 1448, this number is exactly 19.6 %, which is identical to sum of the two component values (Two mapping distance values)
- In practice, sum of two shorter distances is sufficient to provide reliable measure of longer distance.

Interference

- Knowing the existance of double crossover, we can ask, are the crossovers in adjacent chromosome regions independent events or does a crossover region affect the likelihood of there being a crossover in an adjacent region ?
- Crossovers inhibit each other somewhat in an interaction called **interference**.
- If the crossovers in the two regions are independent, we can use the product rule to predict the frequency of double recombinants: that frequency would equal the product of the recombinant frequencies in the adjacent regions.
- In the $v - ct - cv$ recombination data, the $v - ct$ RF value is 0.132 and the $ct - cv$ value is 0.064; so, if there is no interference, double recombinants might be expected at the frequency $0.132 \times 0.064 = 0.0084$ (0.84 percent).
- In the sample of 1448 flies, $0.0084 \times 1448 = 12$ double recombinants are expected.
- But actually only 8 were observed. This suggests some kind of interference.

- Interference is quantified by the term called the **coefficient of coincidence (c.o.c)**,

$$\text{c.o.c} = \frac{\text{Observed frequency (or number) of double recombinants}}{\text{Expected frequency (or number) of double recombinants}}$$

- Interference is defined as:

$$I = 1 - \text{c.o.c}$$

- In our example,

$$I = 1 - \frac{8}{12} = \frac{1}{3} = 33\%$$

- If double recombinants are never observed, c.o.c $= 0$, and so $I = 1$ (complete interference).

Problem

An individual heterozygous for four genes; Aa.Bb.Cc.Dd is test crossed to aabbccdd and 1000 progeny were classified by the gametic contribution of the heterozygous parent as follows:

1. Which genes are linked ?
2. If two pure breeding lines had been crossed to produce the heterozygous individual, what would their genotypes have been ?
3. Draw a linkage map of the linked genes showing the correct genes order and their map distances.
4. Calculate CC and I.

Progeny type	Numbers
aBCD	42
Abcd	43
ABCd	140
abcD	145
aBcD	6
AbCd	9
ABcd	305
abCD	310

Solution

Recombinant loci	Recombinants
A_B	0.10
A_C	0.30
B_C	0.37
B_D	0.10
C_D	0.50

Here, since no recombinant types of A_D are observed, we can conclude that these two genes are completely linked. In a situation like this, we can assume that the combination of two genes inherit as a single entity. Note here the recombination frequencies here is misleading, if taken into consideration for mapping the gene distances. Let us assume "E" represents the entity (combination of A and D) – E for aD and e for Ad.

We can rewrite the progeny phenotypes again as:

Progeny type	Numbers
EBC	42
ebc	43
eBC	140
Ebc	145
EBc	6
ebC	9
eBc	305
EbC	310

Now repeating the process of identifying recombinants again.

Recombinant loci	Recombinants
B_C	0.37
E_B	0.10
E_C	0.30

From this we can infer the map distances and map order, easily. B and C lie the furthest with, E lying on the middle.
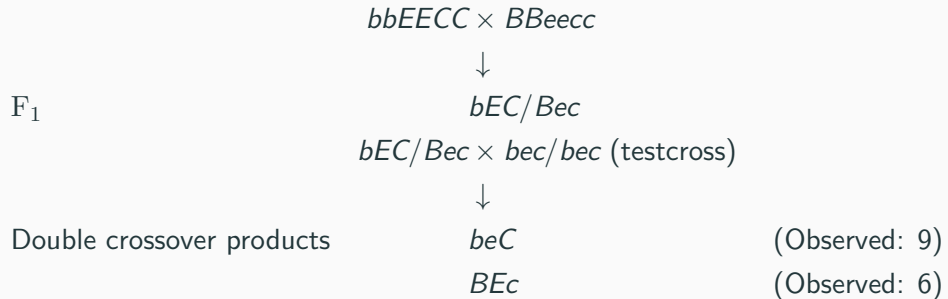
Notice that linear distance between B and C is lesser than the sum of two linear distances. This is because we have not accounted for the possibility of double crossover between these two genes. Let's check whether we can recover the whole map distance accounting for double crossover, by including rarest progeny phenotypes doubly.

Recombinant map
400

Here it is evident that double crossover has infact taken place.

We can calculate the expected frequency of double crossover recombinants, were there no interference from the concerned single crossover events.

Hence at least 30 double recombinants would be expected in 1000 from the given testcross. However, we can note that from the F1 hybrid undergoing following meiosis event, following double recombinants are possible and their observed frequencies in the testcross is:

$$bbEECC \times BBeecc$$

$$\downarrow$$

F_1 *bEC/Bec*

$$bEC/Bec \times bec/bec \text{ (testcross)}$$

$$\downarrow$$

| Double crossover products | *beC* | (Observed: 9) |
| | *BEc* | (Observed: 6) |

Coefficient of coincidence $(CoC) = \frac{\text{Observed double crossover frequency}}{\text{Expected double crossover frequency}}$

Interference $= 1 - CoC = 0.5\%$

Hence the Interference(I) $= 50\%$. Which means that adjacent single crossover event prevents double crossover to occur by 50%.

Bibliography