Linkage, genetic maps

Macular degeneration is a group of diseases characterized by a breakdown of the macula. The macula is the center portion of the retina that makes central vision and visual acuity possible.

"Age-related maculopathy (ARM), also known as age-related macular degeneration (AMD), is the leading cause of irreversible vision loss in the elderly population in the USA and the Western world and a major public health issue. Affecting nearly 9% of the population over the age of 65, ARM becomes increasingly prevalent with age such that by age 75 and older nearly 28% of individuals are affected (1–6). As the proportion of the elderly in our population increases, the public health impact of ARM will become even more severe. Currently there is little that can be done to prevent or slow the progression of ARM (7)."

http://hmg.oupjournals.org/cgi/content/full/9/9/1329#DDD140TB1

Hmmmmm

"It was not long from the time that Mendel's work was rediscovered that new anomalous ratio began appearing. One such experiment was performed by Bateson and Punnett with sweet peas. They performed a typical dihybrid cross between one pure line with purple flowers and long pollen grains and a second pure line with red flowers and round pollen grains. Because they knew that purple flowers and long pollen grains were both dominant, they expected a typical 9:3:3:1 ratio when the F1 plants were crossed. The table shows the ratios that they observed. Specifically, the two parental classes, purple, long and red, round, were overrepresented in the progeny."

http://www.ndsu.edu/instruct/mcclean/plsc431/linkage/linkage1.htm

"Coupling" and "repulsion"

	Observed	**Expected**
Purple, long (*P_L_*)	284	215
Purple, round (*P_ll*)	21	71
Red, long (*ppL_*)	21	71
Red, round (*ppll*)	55	24
Total	**381**	**381**

http://www.ndsu.edu/instruct/mcclean/plsc431/linkage/linkage1.htm

The facts on which Bateson bases his interpretation may be briefly stated in his own words, namely: "that if *A, a* and *B, b* are two allelomorphic pairs subject to coupling and repulsion, the factors *A* and *B* will repel each other in the gametogenesis of the double heterozygote resulting from the union $Ab \times aB$, but will be coupled in the gametogenesis of the double heterozygote resulting from the union $AB \times ab$," and further, "We have as yet no probable surmise to offer as to the essential nature of this distinction, and all that can yet be said is that in these special cases the distribution of the characters in the heterozygote is affected by the distribution in the original pure parents." Bateson further

Tests of significance

The χ^2 test of
"goodness of fit"
(Karl Pearson)

Classical problem

"No one can tell which way a penny will fall, but we expect the proportions of heads and tails after a large number of spins to be nearly equal. An experiment to demonstrate this point was performed by Kerrich while he was interned in Denmark during the last war. He tossed a coin 10,000 times and obtained altogether 5,067 heads and 4,933 tails."

MG Bulmer *Principles of Statistics*

Hypothesis vs. observation

Hypothesis: the probability of getting a tail is 0.5.
Observation: 4,933 out of 10,000.

Well?!!

How can we meaningfully – quantitatively – construct a test that would tell us, whether the hypothesis is, most likely, correct, and the deviation is due to chance – or (alternatively) – the hypothesis is incorrect, and the coin dislikes showing its "head" side for some mysterious reason?

Sampling errors are inevitable, and deviations from perfection are observed all the time.

The goodness of fit test has been devised to tell us, how often the deviation we have observed could have taken place solely due to chance.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

The procedure

Come up with an explanation for the data ("the null hypothesis").

Ask yourself – if that explanation were correct, what should the data have been? E.g., if the hypothesis is that the probability of getting "tails" is 50%, then there should have been 5,000 tails and 5,000 heads. This set of numbers forms the "expected data."

Take the actual – observed – data (**critical point**: take the primary numbers, not the frequencies or percentages – this is because the "goodness of fit" is a function of the absolute values under study).

Plug them into the following formula: $\chi^2 = \sum \frac{(O-E)^2}{E}$

TABLE 5.1 Critical Chi Square Values

Degrees of Freedom	*p* Values: Cannot Reject the Null Hypothesis				Null Hypothesis Rejected		
	0.99	0.90	0.50	0.10	0.05	0.01	0.001
	χ^2 calculations						
1	—	0.02	.45	2.71	3.84	6.64	10.83
2	0.02	0.21	1.39	4.61	5.99	9.21	13.82
3	0.11	0.58	2.37	6.25	7.81	11.35	16.27
4	0.30	1.06	3.36	7.78	9.49	13.28	18.47
5	0.55	1.61	4.35	9.24	11.07	15.09	20.52

χ^2 values that lie in the yellow-shaded region of this table allow you to reject the null hypothesis with > 95% confidence, and for recombination experiments, to postulate linkage.

Calculate *p* value.
If it's .05 or below, the hypothesis is incorrect – the deviation you see in the data is unlikely to be due to chance.
If it's above .05, the hypothesis stands.

SMI?

Take a pure-breeding agouti mouse and cross it to a pure-breeding white mouse. Get 16 children: all agouti (8 males, 8 females). Cross each male with one female (randomly).

Get 240 children in F2: 175 agouti and 65 white (ratio: 2.692).

Calculating the chi square value

Let's hypothesize that we are dealing with simple Mendelian inheritance (the *null hypothesis*). If this were true, then we would ***expect*** that the 240 children would have split: 180 agouti : 60 white.

For agouti mice:

$$(175\text{-}180)^2/180=0.139$$

For white mice:

$$(65\text{-}60)^2/60=0.417$$

sum (Σ) of agouti and white = 0.139 + 0.417 = 0.556

Evaluating the null hypothesis

There are only two classes here, so we must use the "1 degree of freedom" line in the table. For χ^2=0.556, the *p* lies between 0.1 and 0.5.

Our data deviate from the 3 :1 ratio. Statistics tells us, however, that the deviation we saw (not 60, but 65, and not 180, but 175) is observed simply *based on chance* betwen 10% and 50% of the time. This is acceptable: only those deviations that are expected to occur 5% of the time (once every 20 times we do the experiment) or less can force us to say that the deviation is *not* due to chance →

simple Mendelian inheritance for these two alleles

"End of Drug Trial Is a Big Loss for Pfizer" Dec. 4 2006

The news came to Pfizer's chief scientist, Dr. John L. LaMattina, as he was showering at 7 a.m. Saturday: the company's most promising experimental drug, intended to treat heart disease, actually caused an increase in deaths and heart problems. Eighty-two people had died so far in a clinical trial, versus 51 people in the same trial who had not taken it.

Within hours, Pfizer, the world's largest drug maker, told more than 100 trial investigators to stop giving patients the drug, called torcetrapib. Shortly after 9 p.m. Saturday, Pfizer announced that it had pulled the plug on the medicine entirely, turning the company's nearly $1 billion investment in it into a total loss.

The abrupt decision to discontinue torcetrapib was a shocking disappointment for Pfizer and for people who suffer from heart disease. The drug, which has been in development since the early 1990s, raises so-called good cholesterol, and cardiologists had hoped it would reduce the buildup of plaques in blood vessels that can cause heart attacks. Just last Thursday, Pfizer's chief executive, Jeffrey B. Kindler, said publicly that the drug could be among the most important new developments for heart disease in decades and that the company hoped to get Food and Drug Administration approval for it in 2007.

"I'm terribly disappointed," said Dr. Steven E. Nissen, chairman of cardiovascular medicine at the Cleveland Clinic and lead investigator of an earlier torcetrapib clinical trial. "This drug, if it worked, would probably have been the largest-selling pharmaceutical in history."

Sample	Observed	Expected (if drug is harmless)	(O-E)^2	(O-E)^2 div by E	Chi square value
Torcetrapib + lipitor	82	66.5	240.3	3.6	7.23
Lipitor alone	51	66.5	240.3	3.6	

Null hypothesis: torcetrapib is safe (as far as death from cardiovascular events are concerned).
What is the likelihood that the observed difference is due solely to chance?
Somewhere between 0.1 and 1%
The null hypothesis is rejected.

TABLE 5.1 Critical Chi Square Values

Degrees of Freedom	*p* Values: Cannot Reject the Null Hypothesis				Null Hypothesis Rejected		
	0.99	0.90	0.50	0.10	0.05	0.01	0.001
	χ^2 calculations						
1	—	0.02	.45	2.71	3.84	6.64	10.83
2	0.02	0.21	1.39	4.61	5.99	9.21	13.82
3	0.11	0.58	2.37	6.25	7.81	11.35	16.27
4	0.30	1.06	3.36	7.78	9.49	13.28	18.47
5	0.55	1.61	4.35	9.24	11.07	15.09	20.52

χ^2 values that lie in the yellow-shaded region of this table allow you to reject the null hypothesis with > 95% confidence, and for recombination experiments, to postulate linkage.

Back to Bateson and Punnett

Sample	Observed	Expected (if SMI)	(O-E)^2	(O-E)^2 div by E	Chi square value
Purple, long (*P_L_*)	284	215	4761.0	22.1	132.61
Purple, round (*P_ll*)	21	71	2500.0	35.2	
Red, long (*ppL_*)	21	71	2500.0	35.2	
Red, round (*ppll*)	55	24	961.0	40.0	

Null hypothesis: the genes exhibit SMI.
What is the likelihood that the observed difference is due solely to chance?
Well below 0.1%.
The null hypothesis is rejected.

What is going on? What can explain this "repulsion and coupling"?
Why are these two genes disobeying Mendel's second law?

Morgan's observation of linkage

One of these genes affects eye color (*pr*, purple, and pr^+, red), and the other affects wing length (*vg*, vestigial, and vg^+, normal). The wild-type alleles of both genes are dominant. Morgan crossed *pr*/*pr* · *vg*/*vg* flies with $pr^+/pr^+ \cdot vg^+/vg^+$ and then testcrossed the doubly heterozygous F_1 females: $pr^+/pr \cdot vg^+/vg \times pr/pr \cdot vg/vg$.

The data

$pr^+ \cdot vg^+$	1339
$pr \cdot vg$	1195
$pr^+ \cdot vg$	151
$pr \cdot vg^+$	154
	2839

1:1:1:1?! ☺

Sample	Observed	Expected (if drug is harmless)	(O-E)^2	(O-E)^2 div by E	Chi square
AA	1339	710	395641	557.2	1764.1
ab	1195	710	235225	331.3	
Ab	151	710	312481	440.1	
aB	154	710	309136	435.4	

Null hypothesis: genes not linked.
What is the likelihood that the observed difference is due solely to chance?
Ummmmm. Yeah ….
--> null hypothesis, shmull hypothesis.

TABLE 5.1 Critical Chi Square Values

Degrees of Freedom	p Values: Cannot Reject the Null Hypothesis				Null Hypothesis Rejected		
	0.99	0.90	0.50	0.10	0.05	0.01	0.001
	χ^2 calculations						
1	—	0.02	.45	2.71	3.84	6.64	10.83
2	0.02	0.21	1.39	4.61	5.99	9.21	13.82
3	0.11	0.58	2.37	6.25	7.81	11.35	16.27
4	0.30	1.06	3.36	7.78	9.49	13.28	18.47
5	0.55	1.61	4.35	9.24	11.07	15.09	20.52

χ^2 values that lie in the yellow-shaded region of this table allow you to reject the null hypothesis with > 95% confidence, and for recombination experiments, to postulate linkage.

In place of attractions, repulsions and orders of precedence, and the elaborate systems of coupling, I venture to suggest a comparatively simple explanation based on results of inheritance of eye color, body color, wing mutations and the sex factor for femaleness in *Drosophila*. If the materials that represent these factors are contained in the chromosomes, and if those factors that "couple" be near together in a linear series, then when the parental pairs (in the heterozygote) conjugate like regions will stand opposed. There is good evidence to support the view that during the strepsinema stage homologous chromosomes twist around each other, but when the chromosomes separate (split) the split is in a single plane, as maintained by Janssens. In consequence, the original materials will, for short distances, be more likely to fall on the same side of the split, while remoter regions will be as likely to fall on the same side as the last, as on the opposite side. In consequence, we find coupling in certain characters, and little or no evidence at all of coupling in other characters; the difference depending on the linear distance apart of the chromosomal materials that represent the factors. Such an explanation will account for all of the many phenomena that I have observed and will explain equally, I think, the other cases so far described. The results are a simple mechanical result of the location of the materials in the chromosomes, and of the method of union of homologous chromosomes, and the proportions that result are not so much the expression of a numerical system as of the relative location of the factors in the chromosomes. *Instead of random segregation in Mendel's sense we find "associations of factors" that are located near together in the chromosomes. Cytology furnishes the mechanism that the experimental evidence demands.*

T. H. MORGAN

September 10, 1911.

Morgan *Science* 1911

Batrachoseps attenuatus
California Slender Salamander

F.A. Janssens

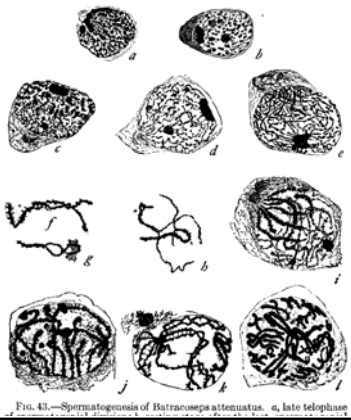
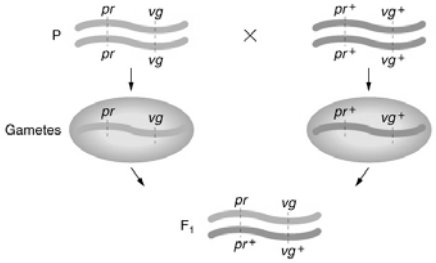


FIG. 43.—Spermatogenesis of Batracoseps attenuatus. *a*, late telophase

These two loci do not follow Mendel's second law because they are linked



The data

$pr^+ \cdot vg^+$	1339
$pr \cdot vg$	1195
$pr^+ \cdot vg$	151
$pr \cdot vg^+$	154
	2839

?

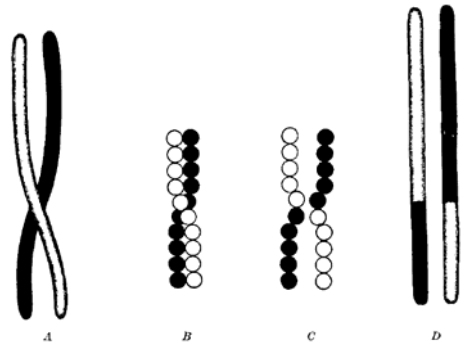
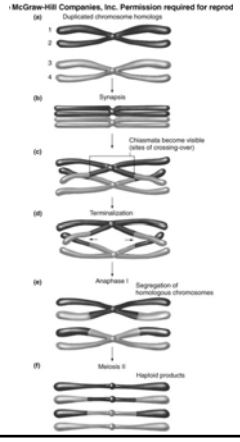
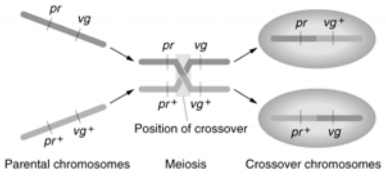
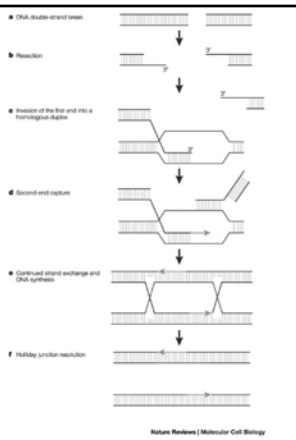
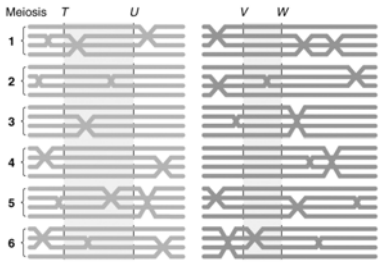


FIG. 24.—Diagram to represent crossing over. At the level where the black and the white rod cross in *A*, they fuse and unite as shown in *D*. The details of the crossing over are shown in *B* and *C*.









Recombination Frequency (Morgan's data)

1339 red, normal
1195 vermillion, vestigial
151 red, vestigial
154 vermillion, normal

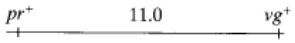
2839 total progeny.
305 recombinant individuals.

305 / 2839 = 0.107

Recombination frequency is 10%.

Recombination frequency → a genetic map (Sturtevant's data)

pr vg/pr vg	165	parental
pr⁺ vg⁺/pr vg	191	
pr vg⁺/pr vg	23	recombinant
pr⁺ vg/pr vg	21	
	400	



Unit definition

1% recombinant progeny =
1 map unit =
1 centimorgan (cM) ~ 1 Mb
(note: the latter applies to humans)

Mapping By Recombination Frequency (Morgan's data)

1339 red, normal
1195 vermillion, vestigial
151 red, vestigial
154 vermillion, normal

2839 total progeny.
305 recombinant individuals.

305 / 2839 = 0.107

Recombination frequency is 10%.

Map distance between the two loci is 10 m.u.

arrangement to be linear and in the group of genes most exhaustively studied, that of the 'sex chromosome' has represented them in a 'chromosome map, as shown in Diagram I.

That the arrangement of the genes within a linkage system is strictly linear seems for a variety of reasons doubtful. It is doubtful, for example, whether an elaborate organic molecule ever has a simple string-like form. Let us, therefore, examine briefly the evidence for or against the idea of linear arrangement of the genes. It is supposed by Morgan that two genes lying in the same chromosome show close linkage if they lie close together, but less linkage if they lie farther apart, and that the farther apart they are the less will be their linkage. As a measure of the distance apart of two genes he

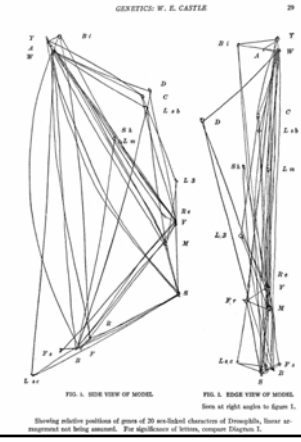
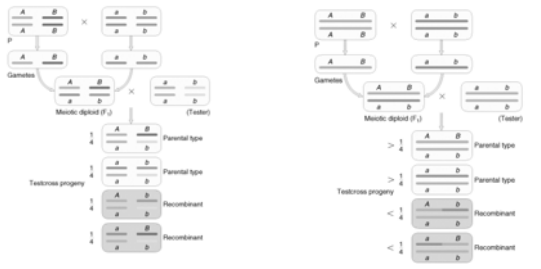


FIG. 1. SIDE VIEW OF MODEL.

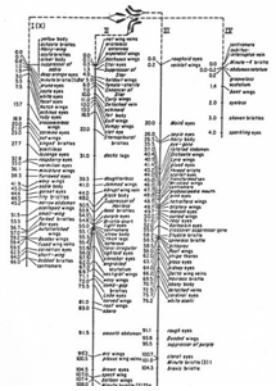
FIG. 2. EDGE VIEW OF MODEL. Seen at right angles to figure 1.

Showing relative positions of genes of 20 sex-linked characters of Drosophila, linear arrangement not being assumed. For significance of letters, compare Diagram I.



If genes are more than 50 map units apart, they behave as if they were unlinked.

The chromosome as a “linkage group”



Bridges (left) and Sturtevant in 1920.

G. Rubin and E. Lewis *Science* **287:** *2216.*

There was an atmosphere of excitement in the laboratory, and a great deal of discussion and argument about each new result as the work rapidly developed.

In 1909 Castle published diagrams to show the interrelations of genes affecting the color of rabbits. It seems possible now that these diagrams were intended to represent developmental interactions, but they were taken (at Columbia) as an attempt to show the spatial relations in the nucleus. In the latter part of 1911, in conversation with Morgan about this attempt—which we agreed had nothing in its favor—I suddenly realized that the variations in strength of linkage, already attributed by Morgan to differences in the spatial separation of the genes, offered the possibility of determining sequences in the linear dimension of a chromosome. I went home and spent most of the night (to the neglect of my undergraduate homework) in producing the first chromosome map, which included the sex-linked genes *y*, *w*, *v*, *m*, and *r*, in the order and approximately the relative spacing that they still appear on the standard maps (Sturtevant, 1913).

Sturtevant 1961

The three-point testcross

From my perspective, the single most majestic epistemological accomplishment of “classical” genetics

Sturtevant, A. H. 1913. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology*, 14: 43-59.

THE LINEAR ARRANGEMENT OF SIX SEX-LINKED FACTORS IN DROSOPHILA, AS SHOWN BY THEIR MODE OF ASSOCIATION

A. H. STURTEVANT

HISTORICAL

The parallel between the behavior of the chromosomes in reduction and that of Mendelian factors in segregation was first pointed out by Sutton (1902) though earlier in the same year Boveri (1902) had referred to a possible connection. In this paper and others Boveri brought forward considerable evidence from the field of experimental embryology indicating that the chromosomes play an important role in development and inheritance. The first attempt at connecting any given somatic character with a definite chromosome came with McClung's (1902) suggestion that the accessory chromosome is a sex-determiner. Stevens (1905) and Wilson (1905) verified this by showing that in numerous forms there is a sex chromosome, present in all the eggs and in the female-producing sperm, but absent, or represented by a smaller homologue, in the male-producing sperm. A further step was made when Morgan (1910) showed that the factor for color in the eyes of the fly *Drosophila ampelophila* follows the distribution of the sex chromosome already found in the same species by Stevens (1908). Later, on the appearance of a sex-linked wing mutation in *Drosophila*, Morgan (1910a, 1911) was able to make clear a new point. By crossing white-eyed, long-winged flies to those with red eyes and rudimentary wings (the new sex-linked character) he obtained, in F_2, white-eyed,

Reading

Two chapters from Morgan's book (III, on linkage, and V, on chromosomes).

A short chapter from Sturtevant's *History of Genetics*.

Chapter 5, section 2.

How to Map Genes Using a Three-Point Testcross

1. Cross two pure lines.
2. Obtain large number of progeny from F1.
3. Testcross to homozygous recessive tester.
4. Analyze large number of progeny from F2.

P $v+/v+ \cdot cv/cv \cdot ct/ct \times v/v \cdot cv+/cv+ \cdot ct+/ct+.$

↓

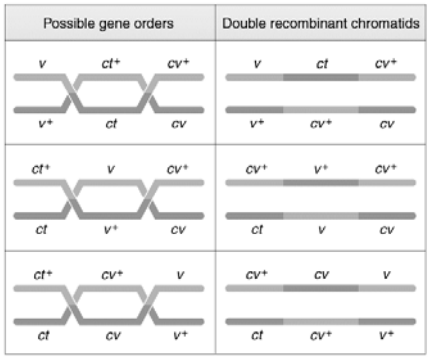
F1 $v/v+ \cdot cv/cv+ \cdot ct/ct+ \times v/v \cdot cv/cv \cdot ct/ct.$

↓

Two *Drosophila* were mated: a red-eyed fly that lacked a cross-vein on the wings and had snipped wing edges to a vermilion-eyed, normally veined fly with regular wings. All the progeny were wild type. These were testcrossed to a fly with vermilion eyes, no cross-vein and snipped wings. 1448 progeny in 8 phenotypic classes were observed.

Map the genes.

$v \cdot cv^+ \cdot ct^+$	580
$v^+ \cdot cv \cdot ct$	592
$v \cdot cv \cdot ct^+$	45
$v^+ \cdot cv^+ \cdot ct$	40
$v \cdot cv \cdot ct$	89
$v^+ \cdot cv^+ \cdot ct^+$	94
$v \cdot cv^+ \cdot ct$	3
$v^+ \cdot cv \cdot ct^+$	5
	1448



1. Rename and rewrite cross

For data like these, no need to calculate χ^2. Begin (you don't *have* to, but it helps) by designating the genes with letters that look different in UPPER and lowercase (e.g., not "W/w" but "Q/q" or "I/i"):

eye color: v^+/v = E/e
vein on wings: cv^+/cv = N/n
shape of wing: ct^+/ct = F/f (you fly using wings)

```
P:            EE nn ff   x   ee NN FF
test-cross:   Ee Nn Ff   x   ee nn ff
```


2. Rewrite data

Arrange in descending order, by frequency.

NCOs	e	N	F	580
	E	n	f	592
	e	n	F	45
	E	N	f	40
	e	n	f	89
	E	N	F	94
DCOs	E	n	F	5
	e	N	f	3

3. Determine gene order

e	N	F	580
E	n	f	592

With the confusion cleared away, determine gene order by comparing most abundant classes (non-recombinant, NCO) with double-recombinant (least abundant, DCO), and figuring out, which *one allele pair needs to be swapped between the parental chromosomes in order to get the DCO configuration.* This one allele pair will be of the gene that is in the middle.

E	n	F	5
e	N	f	3

3b. Determine gene order

NCOs: DCOs:

Enf → EnF

eNF eNf

Gene order: E F N (or N F E).

4. E and F

Next, map distance between genes E and F by comparing the number of single recombinants (COs) for those two genes with the number of NCOs.

e	N	F	580
E	n	f	592
e	n	F	45
E	N	f	40
e	n	f	89
E	N	F	94
e	N	f	3
E	n	F	5

RF=(89+94+3+5)/1448=0.132

The E and F genes are separated by 13.2 m.u.

4b. F and N

Now, map distance between genes F and N by comparing the number of single recombinants (COs) for those two genes with the number of NCOs.

e	N	F	580
E	n	f	592
e	n	F	45
E	N	f	40
e	n	f	89
E	N	F	94
e	N	f	3
E	n	F	5

RF=(45+40+3+5)/1448=0.064

The F and N genes are separated by 6.4 m.u.

4c. E and N

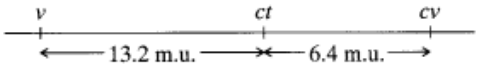
Finally, map distance between genes E and N by comparing the number of single recombinants (COs) for those two genes and the number of DCOs for those two genes with the number of NCOs. Count DCOs twice because they represent *two* recombination events, and to calculate the correct RF we must, by definition, count every recombination event that occurred between those two genes (even if it doesn't result in a recombinant genotype for those two genes!).

e	N	F	580
E	n	f	592
e	n	F	45
E	N	f	40
e	n	f	89
E	N	F	94
e	N	f	3
E	n	F	5

RF=(45+40+89+94+3+5+3+5)/1448=0.196

The E and N genes are separated by 19.6 m.u.

5. The map (ta-daaa!)



6. Interference

A crossover event decreases the likelihood of another crossover event occurring nearby.

$$1 - \left[\frac{\text{observed frequency, or number of double recombinants}}{\text{expected frequency, or number of double recombinants}} \right]$$

Final map:

E ——————————F———N
13.2 m.u. 6.4 m.u.
|---------- 19.6 m.u.----------|

For dessert, do not forget to calculate interference for these loci. The mathematical probability of seeing a DCO in this area is equal to the product of probabilities of seeing a CO between E--F and seeing a CO between F--N:

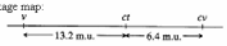
p(expected DCOs)=0.132 x 0.064=0.008448

This means we should have seen 0.008448 x 1448= 12 DCOs. We only saw 3 + 5 = 8, i.e. the observed frequency of DCOs is 8/1448 = 0.005524.

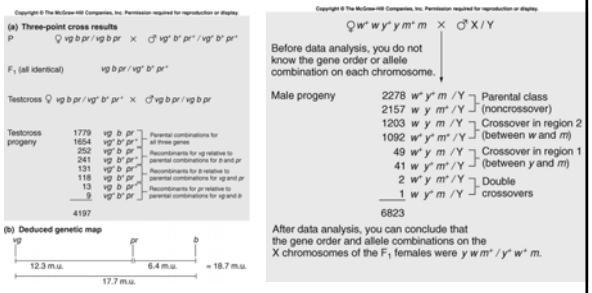
Interference is equal to 1 minus the "coefficient of coincidence" = 1 - p(O)/p(E) = 35% → 35% of the double-recombination events that were expected to have occurred based on probabilistic considerations didn't because of interference.

Recombination analysis relies so heavily on three-point testcrosses and extended versions of them that it is worth making a step-by-step summary of the analysis, ending with an interference calculation. We shall use numerical values from the data on the *v*, *ct*, and *cv* loci.

1. Calculate recombinant frequencies for each pair of genes:
   v–*cv* = 18.5%
   cv–*ct* = 6.4%
   ct–*v* = 13.2%
2. Represent linkage relations in a linkage map:



3. Determine the double recombinant classes.
4. Calculate the frequency and number of double recombinants expected if there is no interference:
   Expected frequency = 0.132 × 0.064 = 0.0084
   Expected number = 0.0084 × 1448 = 12
5. Calculate interference:
   Observed number of double recombinants = 8
   Expected number of double recombinants = 12
   ∴I = $1 - \frac{8}{12} = \frac{4}{12}$ = 0.33, or 33%

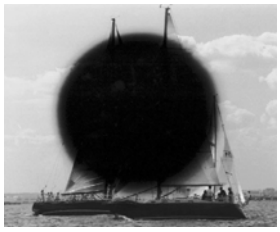


Mapping by linkage

Two SNPs showed the greatest linkage, and they lie in a 260 kb region. This stretch contains the complement H gene – CFH is a component of the innate immune system which regulates inflammation, which, in turn, is consistently implicated in AMD.

"Resequencing revealed a polymorphism in linkage disequilibrium with the risk allele representing a tyrosine-histidine change at amino acid 402. This polymorphism is in a region of CFH that binds heparin and C-reactive protein. Individuals homozygous for the risk alleles have a 7.4-fold increased likelihood of AMD (95% CI 2.9 to 19)."

Haines et al. *Science* **308**: 419.



Daiger *Science* 308: 362.

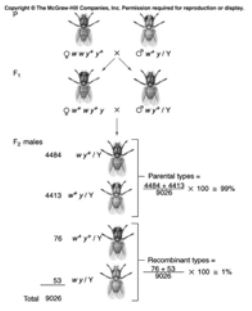


Fig. 5.2