

## C2005/F2401 '10 Lecture #23

© Copyright 2010 Deborah Mowshowitz and Lawrence Chasin Department of Biological Sciences Columbia University New York, NY  
Last Updated: 12/08/2010 02:37 PM

Handouts: [23A](#) = Crossing Over & Recombination Frequency  
[23B](#) = Hardy Weinberg Law  
23C = How to measure RF's in diploids. This handout is on the web only -- it was not provided in class.  
[http://www.columbia.edu/cu/biology/courses/c2005/2n\\_rf07.html](http://www.columbia.edu/cu/biology/courses/c2005/2n_rf07.html)

### I. Independent Assortment & Linkage

#### A. How do you distinguish linkage and independent assortment?

1. *What you do*: You measure the proportion of parentals and recombinants from **many** meioses of a double heterozygote (AB/ab or Ab/aB).

#### 2. Definition

a. **Linkage** = if you get more parentals than recombinants from many meioses.

b. **Independent assortment** = if you get equal numbers of parentals and recombinants from many meioses. (You can't get more recombinants than parentals from the pooled results of **many** meioses. See 'important note' below.) There are two ways to get independent assortment:

- (1). Genes are on separate chromosomes
- (2). Genes are very far apart on the same chromosome.

Details: If genes alpha and beta are on the same chromosome, but relatively far apart, then meioses can occur with multiple crossover events (not shown on handout 22B). If alpha and beta are far enough apart, they can be switched back and forth multiple times, so that on the average they are equally likely to end up switched or not. So if alpha and beta are very far apart on the same chromosome they will act unlinked -- that is, they will assort independently = you will get all 4 kinds of gametes in equal proportions. (Same result as if the genes were on separate chromosomes.) More on this below.

#### B. How is frequency of crossing over (recombination) related to distance?

- Frequency of actual crossing over (between any two points on the DNA) is always proportional to the distance between the points
- Crossing over occurs relatively frequently. However, the chance a given crossing over event will fall between any two points depends on the distance between the two points.
- If two genes (or mutations or 'markers') are far apart, the chance of a crossover landing between them and producing recombinant chromosomes is large.
- If two genes are close together, the chance of a crossover between them is small.
- Therefore the **real** frequency of crossing over between any two genes (or mutations) can be used as a measure of distance between the 2 genes (or the two mutations).
- The **measurable** frequency of crossing over (RF) is only proportional to distance up to a point, because of complications due to multiple crossovers, as explained below.
- A high frequency of crossing over corresponds to a low degree of linkage and vice versa.

#### C. How do you measure the extent of linkage between genes A and B?

1. *Recombination Frequency (RF) is used to measure the frequency of crossing over.*

RF =  $\frac{\text{Recombination Frequency}}{\text{Recombination Frequency}}$  = % of haploid meiotic products (gametes or spores) that are recombinant  
= (# recombinants/total products) X 100  
= frequency of recombinants that are recovered

2. *Why use RF?* RF is calculated by examining the products of many meioses, not one. RF is used as an indication of the actual incidence of crossing over because we seldom examine the results of a single meiosis. Instead, we look at the total results from many meioses.

3. *RF is proportional to distance within the proper range*

- What is the 'proper range'? When there is either zero or one crossover per meiosis (in the interval being checked). In this case, what looks like a parental is parental (no crossovers) and what looks like a recombinant is

recombinant (result of a crossover). In this range, RF and distance are proportional.

- What is outside the proper range? When there are multiple crossovers per meiosis (in the interval being checked). In this case, you 'lose' some recombinants -- multiple crossovers can put the genes back into the parental combination, and what looks like a parental may be a multiple recombinant. In this range, RF does not reflect the real incidence of crossing over, and RF and distance are not proportional, as explained below.

## II. Understanding RF -- Relationship of RF to Distance. Sections A-C will be discussed in class in parallel.

**A. How will an individual meiosis go?** Need to consider results of individual meioses to see what results (values of RF) to expect from multiple meioses. For outcomes of individual meioses, see handout 22B, left panel titled "Crossing over/ Linkage" and/or 23A. Suppose you start with parental chromosomes at meiosis (say AB/ab). What haploid products (gametes or spores) will you get from a single meiosis? We've already considered two possibilities:

1. **Type 1 -- No Crossovers.** If there is no crossover in a meiosis you get all parental products. See (1) on handout 22B or 23A or Becker 20-15 (20-16), case (b).

2. **Type 2 -- One Crossover.** If there is one crossover event in a meiosis you get the 1/2 parental, 1/2 recombinant products. See (2) on handout 22B or 23A or Becker 20-15 (20-16) case (c), or 20-16 (20-17).

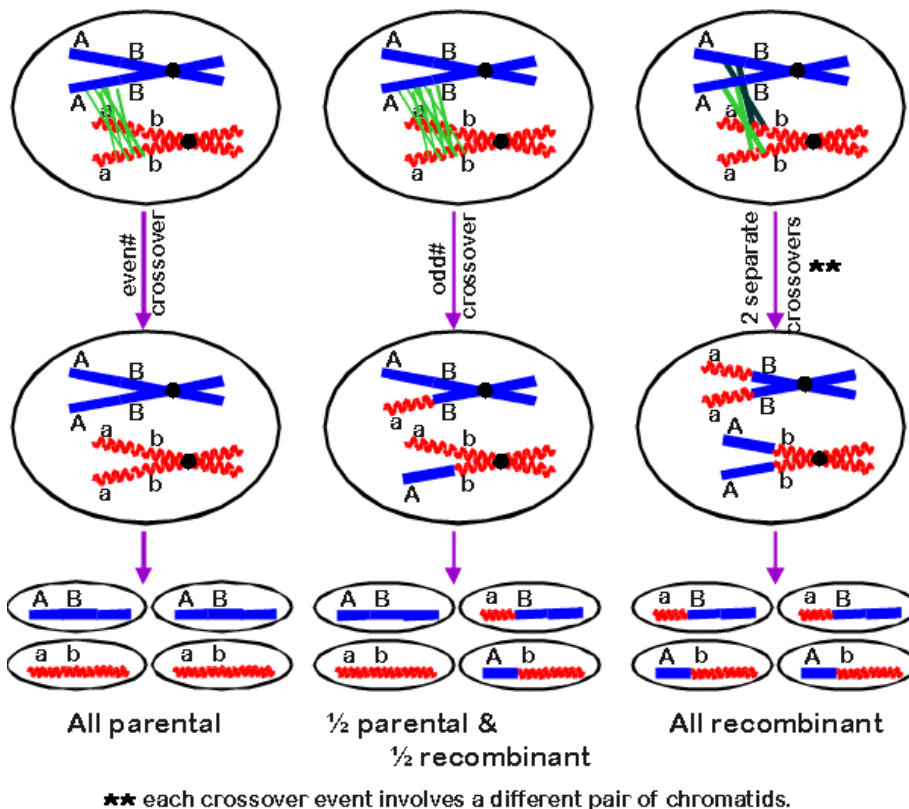
There is a third possibility shown on handout 23A, but not on handout 22B:

3. **Type 3 -- Multiple Crossovers.** If there are multiple crossovers in one meiosis, there are several possible outcomes. Multiple crossovers (in any one meiosis) can give either all parental haploids, or all recombinant haploids, or 50 - 50. See (3) on handout 23A. What happens in an individual meiosis depends on whether the total # of crossovers is even or odd and which chromatids are involved. For more details, see 'Multiple Crossovers' below. **This diagram is included FYI only.**

**Important note:** The diagrams referred to above show the possible results of any **one** individual meiosis. If you look at the pooled results of **many** meioses, you **never** get all recombinant products -- you never get more than 50% recombinants. Type 1 meioses don't give any recombinants, type 2 meioses always give 1/2 recombinants, and the pooled result of many type 3 meioses is 1/2 recombinants and 1/2 parentals. So overall, if there are many crossovers, you get 1/2 parentals and 1/2 recombinants.

Figure below shows possible alternative ways you can get multiple crossovers -- This figure is included FYI for the experts.

### Multiple Crossovers



**Explanation of figure above:** If A and B are very far apart, the average meiosis will involve multiple crossover events. The resulting

gametes can be recombinant, parental, or a mixture, depending on whether the total # of crossovers is even or odd and whether crossovers involve the same pair of chromatids (crossing over more than once) or more than one pair (a different pair of chromatids for each crossover). Any individual meiosis with multiple crossovers can give you all parental gametes, half parental and half recombinant, or all recombinant gametes, as shown in the 3 cases above. The **average** result of **many** meioses with multiple crossovers is 50/50 parental and recombinant gametes. So if you look at many gametes from many meioses involving multiple crossovers, the total gametes will be 50% parental and 50% recombinant.

## B. What will you get from many meioses? How does RF change with distance?

**1. How will many meioses go?** Suppose 2 genes (or mutations, or 'markers') are on the same chromosome. The chart below summarizes the correlation between RF, distance, type of individual meiosis, and types of gametes from a total of many meioses. The curve below (also on handout 23A) shows how RF changes with distance. Here are some of the possible cases:

**a. If the 2 genes are very close**, almost all meiosis are type 1, and almost all products are parental. As distance between the genes decreases, RF approaches a limit of zero.

**b. If the 2 genes are close, but not as close**, most meioses are type 1, but a few are type 2. Therefore most products are parental, but some are recombinant. RF will be small, but greater than zero.

**c. If the 2 genes are farther apart (than in previous case)**, most meioses are still type 1, but a larger percent are type 2. Therefore most products are parental, but a larger percent (than in the previous case) are recombinant. RF will be bigger than in the previous case. As long as there are few or no meioses of type 3, the RF will be proportional to the distance between the genes. See graph below -- you are in the linear part of the curve. Map distance can be calculated from this part of the curve, as explained below.

**d. If the 2 genes are far enough apart, some meioses will be type 3.** If distance is far enough so there are multiple crossovers, RF will not be proportional to the distance between the genes. You are in the part of the curve that levels off. (See below for why the curve levels off.)

**e. If the 2 genes are very far apart**, almost no meioses are type 1, and virtually all meioses are type 2 or 3. Therefore you will get 1/2 parental and 1/2 recombinant products, and RF will be 50%. Why is this?

- Each type 2 meiosis gives you 1/2 and 1/2, recombinant and parental.
- Each individual type 3 meiosis can give either all parental, all recombinant, or 1/2 and 1/2. The net result of many type 3 meioses will be 1/2 parental and 1/2 recombinant.
- What will RF be? If haploids are 1/2 and 1/2, parental & recombinant, RF = 50%. This is the maximum value of RF. This means that genes far apart on the same chromosome will assort independently just like genes on separate chromosomes [Becker 20-15 (20-16), case (a) or Sadava fig. 12.8 (10.8)]. In both cases, you will get 50% recombinants and 50% parental, or RF = 50 %.

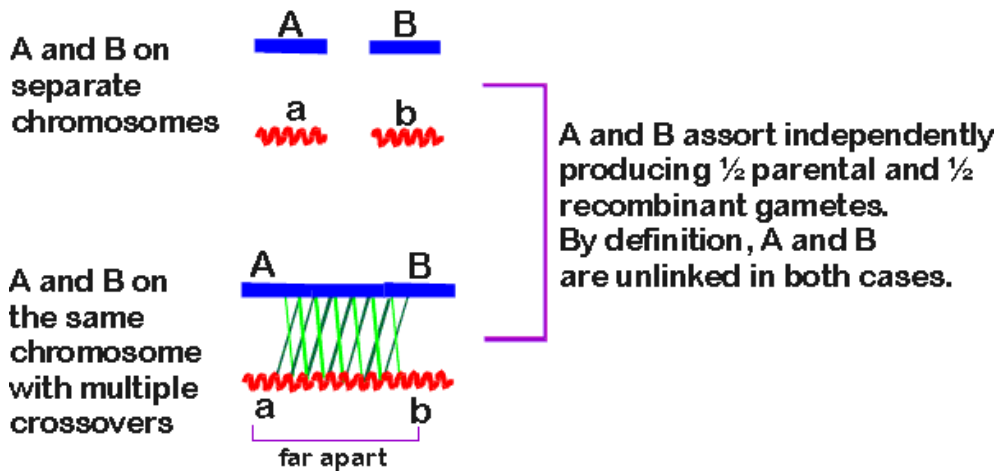
## 2. How can you get independent assortment?

**a. The genes can be on separate chromosomes** -- see top case in picture below, or last lecture.

**b. The genes can be far apart on the same chromosome.** How will this result in independent assortment? Here are two ways to see it:

(1). If genes are far apart, virtually all meiosis are type 2 or 3, so 1/2 gametes are parental and 1/2 recombinant as explained in e above.

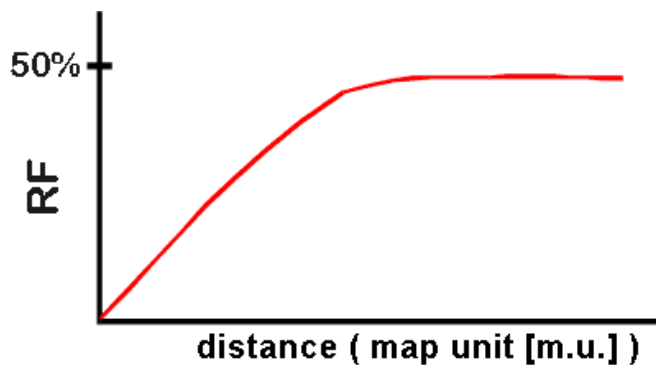
(2). Multiple crossovers will eliminate the linkage -- see bottom case in picture below. Suppose there are multiple crossovers between the genes. An odd number of crossover events will produce a recombinant; an even number of crossovers will switch it back, and produce a parental combination. If there are many crossovers, the number of even crossovers should be about equal to the number of odd crossovers, so the number of parental and recombinant combinations should be about equal. In other words, A is just as likely to end up connected to B or to b.



c. **Physical Linkage (placement on the same chromosome) does not always lead to genetic linkage.** Genes that are not genetically linked can be on the same chromosome, as in b above. Whether genes are considered linked or not depends (by definition) on the ratio of recombinants/parentals. If the ratio is 1 (1/2 recombinants and 1/2 parentals, or RF = 50%) the genes are considered unlinked (genetically), whatever the physical relationship of the two genes. All that matters is whether or not each allele of gene A (A or a), has a 50% chance of ending up in a haploid with either allele of gene B (B or b). How it happens doesn't matter -- if it does, the genes are said to be unlinked.

### C. Summary of Relationship of RF & Map Distance

#### 1. Curve of RF vs. Distance



a. **Map Distance -- using the linear part of the curve.** RF is proportional to map distance in the linear part of the curve. See below for units.

b. **Why does the curve level off as shown?** As A and B get relatively far apart, multiple crossovers start to occur. The number of crossovers increases linearly with distance, but the number of detectable crossovers does not continue to increase linearly. This is because some multiple crossovers switch the A's and B's back to where they were in the first place. Only those crossovers that switch parental alleles to give new allele combinations can be detected and counted as recombinants. Multiple crossovers that switch the alleles back to the parental combination are not counted as recombinants -- they are considered parentals. If you take genetics, you will learn all the ins and outs of counting recombinants and measuring distances, but we will not go beyond this point. However you should note that the max. RF is 50%, not 100%. A single meiosis with several crossovers can produce 100% recombinant gametes, but if you look at the combined results of **many** meioses, you never get more than 50% recombinants overall.

#### 2. Summary Chart-- What Happens When A-B Distance Changes?

Distance	per meiosis	from many meioses	Overall	Linkage
As A-B distance declines	Type 1 -- no crossovers --most often	all parentals	RF → 0	Approaches 100% (complete linkage)
As A-B dist. increases	Type 2 -- 1 crossover --more often	mostly parentals	RF is proportional to distance	Genes Linked (extent depends on distance)
As A-B dist. gets very large	Type 3 --multiple crossovers	50/50	RF levels off at 50%	Approaches none -- genes act unlinked

To review the terminology and significance of crossing over, do problems 10-6 & 10-7.

### III. Mapping & Wrap up of Linkage and Crossing over

#### A. Mapping -- How do you measure and use RF?

1. *Do the Cross*: Cross two double homozygotes to get a heterozygote, and then get heterozygote to go through meiosis and tally products of meiosis.

If you do AAbb X aaBB, heterozygote will be Ab/aB. Gametes will be AB, ab, Ab, and aB.

If you do AABB X aabb, heterozygote will be AB/ab.

If you cross two mutants with mistakes at different points in the DNA, cross will be like this:

Parent 1	X	Parent 2	----->	Heterozygote
---X-----		-----X---		---X-----
---X-----		-----X---		-----X---

What will gametes be this time?

In all these cases, it is important to keep track of what alleles or mutations are on one homolog and what is on the other (in the parents).

2. *Calculate RF*. Once heterozygote goes through meiosis, classify haploid products of meiosis as parental or recombinant and calculate RF using formula as above.

a. **If products of meiosis are spores** -- in this case, haploids can be grown by mitosis and their phenotype (& genotype) directly classified as parental or recombinant.

b. **If products of meiosis are gametes** -- in this case, determining genotypes of products of meiosis cannot be done directly, and you have to look at the diploid organisms that are formed from the gametes. From the phenotypes of the diploid zygotes/organisms you infer the genotypes of the gametes. This is discussed below and in detail on [23C -- How to do RF's with Diploids..](#)

3. *How you make a simple map.*

a. **The principle**: RF is proportional to distance, up to a point, as explained above. Therefore, within the proper range (see below), map distances are additive, just like regular distances.

b. **Units**: 1% RF corresponds to 1 map unit (in the proper range). One map unit is also known as one centiMorgan or 1 cM.

c. **Procedure -- An example**: Suppose you want to order genes A, B and C, and you do the appropriate crosses. For example:

AB/ab --(meiosis)--> 14% Ab and aB  
Bc/bC --(meiosis)--> 4% BC and bc

Then RF between gene A and gene B is 14% and distance is 14 mu or cM; RF between gene B and gene C is 4% and distance between them is 4 mu or cM. Where is gene C? Data put gene C 4 cM from B, but C could be on side nearer to A or away from A. How do you tell which case it is? You need to measure the RF between A and C. It will be 10 or 18%, depending on whether order is A-C-B or A-B-C. For a typical map, see Sadava fig. 12.21 (10.21); for a worked out example try fig. 12.22 (10.22).

4. *Why maps are not completely additive.*

a. **How crossovers are counted**. Double crossovers and no crossovers both → parental allele combinations in the gametes and are counted as "parentals," so RF's don't really count # switches, but approximate it -- RF's really measure the # of **detectable** recombinant combos in products of meiosis. (See legend to graph.)

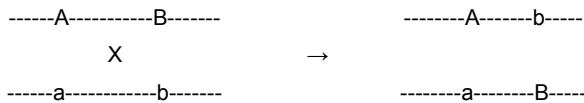
b. **When you can ignore multiple crossovers**. If you stick to low values of RF and distance, in the linear part of the curve, then you can ignore multiple crossovers, since they are rare. In that case, RF and distance are proportional. For most purposes, values under 25% are considered ok. If you use larger values, you have to correct for multiple crossovers. (How to do so will not be discussed here; it will be covered in genetics courses.)

To go over how to relate RF and map distance, try problem 10-11, parts A-D.

To review complementation vs crossing over and get practice making a map, try problem 11-14. (Note that in this problem, crossing over is occurring between mutations -- locations on the DNA as shown in example above -- as vs. between genes A and B.)

**B. When and how does crossing over occur?** Some details to review and/or notice. This is for reference; will not be covered in class.

We sometimes draw crossing over as if there were two single chromosomes (one chromatid per chromosome) involved like so:



But sometimes we draw crossing over as if occurred when each chromosome is already doubled, and there are 2 chromatids per chromosome. (See pictures below.) Which is it?

1. *Each single crossover event involves one pair of chromatids*, but crossing over occurs at a stage (prophase I of meiosis) when there are 4 homologous chromatids, 2 per chromosome.

2. *Crossing over happens only at meiosis (pro. I)*, not mitosis. Only affects next generation, not the generation in which it occurs. If crossing over occurs in the germ cells of a multi-cellular organism, the gametes of the organism are changed, but the somatic cells of the organism are unaffected.

3. *Crossing over requires at least two things*

a. **Enzymes** for pairing, cutting, and rejoining of DNA. Some of the enzymes involved are probably the same ones as for repair of damaged DNA. These are NOT restriction enzymes. The enzymes of restriction/modification and the enzymes of recombination are different.

b. **Homologous DNA sequences.**

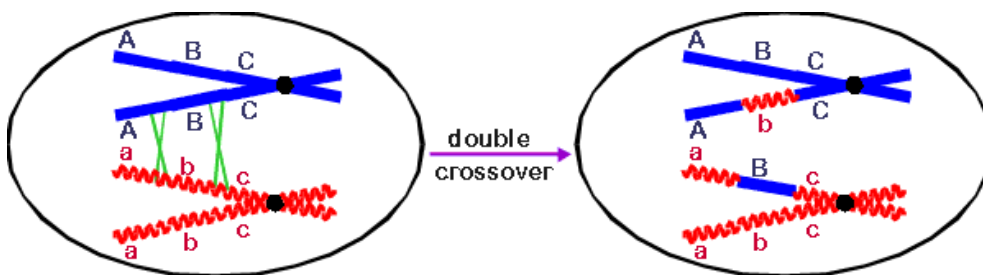
4. *Why does it occur only in meiosis?* Some of the necessary enzymes for pairing, cutting and rejoining are active only in germ cells at pro. I. (Homologous DNA's are present in all 2N cells. However, homologous chromosomes are paired at Prophase I of meiosis, but not at prophase of mitosis.)

5. *Each crossover involves an actual cut and rejoining between two molecules.* This has some consequences.

a. **Crossing over is Reciprocal.** Every time there is a cut and rejoin that produces, say, Ab, a reciprocal aB is also produced.

b. **A double crossover involves 2 separate events.** A double crossover requires two separate cut and rejoining events. In the case shown below, both events involve the same pair of chromatids. If ABC and abc crossover to give AbC and aBc, that's a double crossover -- it takes two cut and rejoining events and is rarer than either one alone.

This has practical consequences. It means that you can usually tell double recombinants from single recombinants, even if you don't know the gene order, by seeing which types of recombinants occur less frequently. This often helps to establish gene order. For example, in this case, since the genes are in order of A-B-C, then ABc and abC recombinants should be more common than AbC and aBc. If the order of genes were A-C-B, then AbC -- really ACb -- would be more common than ABc --really AcB -- and so on.)



To review crossing over, map units and the effects of double crossovers, see problem 10-8. (Note this problem is about a haploid organism with an unfamiliar life cycle. Be sure you know the answer to part A before continuing.)

C. How do you count recombinants in a diploid organism? See handout [23C -- How to do RF's with Diploids](#) (on line).

1. *The problem:* You can't look directly at gametes -- you need to look zygotes (or at diploid organisms that develop from zygotes by mitosis). Then you have to infer the genotype of the gametes from the phenotype of the zygotes.

2. *The solution:* To make the results easier to analyze, you usually make one parent doubly heterozygous and one doubly homozygous recessive. (See note at \*\* below.) In other words, you do a test cross and analyze zygotes. (Test cross = Any cross where one parent is homozygous recessive for all genes under consideration.) Advantages of a this particular set up:

a. **You don't have to worry about crossing over in the homozygous parent.** If it occurs, it has no effect on the gametes.

b. **Homozygous parent can contribute only recessive alleles.**

(1). Any dominant allele came in the zygote came from the heterozygous parent. Therefore you can deduce genotype of zygote from its phenotype.

(2). You can easily classify each zygote as parental or recombinant -- that is, coming from a

parental or recombinant gamete from the heterozygous parent.

3. **Example:** Suppose you do the cross  $Ab/aB \times aabb$ , and you get  $A\_B\_$  pheno offspring. Then

a. **What goes in the blanks?** Has to be a and b.

b. **How was the zygote formed** -- from recombinant or parental gamete from  $Ab/aB$ ? Must be AB gamete from double heterozygote parent met ab from homozygous recessive parent.

4. **More info:** Details of how to measure the RF in a diploid are explained on [Handout 23C = How to measure RF's with diploids](#). For an example of such a cross, see Sadava figs. 12.18 & 12.20 (10.18 & 10.20). These are both about the same cross. For an example of how you use similar results to make a map, see fig. 12.22 (10.22).

\*\*Note: If you did a cross with fruit flies in intro lab, you may have set it up differently because there is no crossing over in male flies.

**For problems measuring RF in diploids, see problems 10-10, 10-11E, and 10-12 to 10-14. (There are additional problems on this in 10 and 10R.)**

## IV. Population Genetics

### A. Intro. We need to shift from "How do things work?" to "How did they get that way?"

1. **The Question:** We know how mutations occur, but why do some spread and others do not? Why is blood type O more common than B? Why is CF commoner in whites, Sickle Cell Disease (SCD) in blacks? Why do genes for drug resistance change, but gene for cytochrome c stays the same! Why is there more variation in introns than in exons? In other words, how did the particular state of affairs that now exists come to be this way?

2. **The short Answer:** How did we get this way? All the available data indicates it happened through evolution by natural selection (= common descent with modifications). Provides answers to all questions in (1). Will fill in some details below.

3. **Importance of Populations.** We need to look at genetics of populations, not just genetics of individuals to answer these questions. For example, how do you get from population of sensitive flies to population of resistant flies? (Mutation does it for one fly, but how does it spread?)

4. **So how do you analyze change in a population?** This is the next topic. First we'll consider what happens when there is no change; then how change occurs.

### B. How to treat Genetics of Populations vs genetics of individuals (for diploids)

1. **Consider a specific case** -- if everybody is  $Aa$ , what is the genotype of the population, and will it stay that way?

a. **What is the genotype of this population** (= genetic structure of population)?

$f(a) = f(A) = 1/2$ ;  $F(Aa) = 1$ ;  $f(aa) = f(AA) = 0$ . [ $f(a)$  means frequency of allele a, and so on.]

b. **What will next ( $F_1$ ) generation be?** Since the only possible cross is  $Aa \times Aa$ , you will get 1:2:1 ( $AA:Aa:aa$ ) in the  $F_1$  generation. So after one generation, genotype of population is:

$f(a) = f(A) = 1/2$ ;  $F(Aa) = 1/2$ ;  $f(aa) = f(AA) = 1/4$ .

c. **What will next ( $F_2$ ) generation be?** Either "pot" reasoning (see below) or considering all possible individual crosses will show you that the  $F_2$  generation will have the same genotype as the  $F_1$  if you assume

(1). Random mating (Choice of a mate is made at random. Therefore chance of a particular cross, say  $AA \times aa$ , depends only on frequency of parental types ( $AA$  and  $aa$ ) in population.)

(2). Large population (There are many matings and many offspring, so gametes and zygotes made are representative of the whole population.)

(3). No selection (Everyone has equal chance of finding a mate, and same average # offspring for each type of cross -- therefore  $AA$ ,  $aa$  and  $Aa$  have an equal chance of passing on their alleles).

For another example of how you reach equilibrium in one generation, see Sadava fig. 21.7 (22.7).

Note that random mating is **not** the same as no selection. (See below for a longer explanation of this.)

- **No selection** = every person is just as likely to find a mate, and has same # descendants on the average = selection of whose **alleles will be passed on** is random (with respect to phenotype/genotype).
- **Random mating** = choice of mate is random. Everybody picks a spouse independently of his or her phenotype/genotype = choice of **which mate** is random (with respect to phenotype/genotype) -- which type you pick is proportional to the frequency of that type in the population. You pick a random sample of what is "out there." (Can have a case where everyone finds a mate, but the choice of mate is not random. That would be a case of no selection, but nonrandom mating. See below.)

**d. What is "pot" reasoning?** Assume all alleles of population are in one "pot" (usually called the gene pool) something like a lottery drum. Can think of pot or drum as full of slips; each slip has one allele written on it. In this case, there are lots of slips and 1/2 have "a", half have "A." Chance an individual will be born of genotype AA is same as chance you will pick two slips with "A" out of the pot (if you have two honest, independent, picks). In this case, chance of picking AA is 1/4, chance of picking aa is 1/4 and chance of picking Aa = 1/2. (1/4 chance of Aa + 1/4 chance of aA)

**2. Genetic equilibrium.** What if you start with 1/2 AA and 1/2 aa? Same as above by F<sub>1</sub>! (If there is random mating, etc.) So call 1:2:1 genetic equilibrium if start with two alleles A, a, and have 1/2 A and 1/2 a. Equilibrium is reached in 1 generation and stays that way. This is the expected result if pot with 1/2 A and 1/2 a is well mixed.

### 3. General case = Hardy-Weinberg Law

**a. What is the H-W law?** It is the general statement (as opposed to the specific example given above) of what proportions (of alleles and genotypes) you should have at equilibrium. Or, if you start with two alleles not necessarily at 1/2 and 1/2, what happens?

Call  $f(A) = p$ ;  $f(a) = q$ . Then  $p + q = 1$  since every allele is either A or a.

Note: p and q do NOT have to be equal to 1/2. They can be any value as long as  $p + q = 1$ .

At equilibrium:

$f(AA) = p^2$ ;  $f(aa) = q^2$ , and  $f(Aa) = 2pq$ , and  $p^2 + 2pq + q^2 = 1$ , since every individual is either AA, aa, or Aa.

#### b. How did we get this?

You can arrive at the H-W law by considering what will happen if you have a big pot full of slips; the fraction with A is p; the fraction with a is q; then consider the chance of picking two slips with A, two with a, etc. as above.

**c. What is the genotype of the population in the general case?** Genotype of Population (aka "genetic structure of population") is described if know  $f(A)$ ,  $f(a)$  and  $f(AA)$ , etc. = allele frequencies and genotype frequencies.

(1). If population is in equilibrium, entire genotype of population follows once you know p and q. If you know frequencies of alleles, you can calculate frequency of genotypes using the H-W law. (Or vice versa.) Some examples are given below.

(2). If population is NOT in equilibrium, you can still define the genotype of the population if you know  $f(A)$ ,  $f(AA)$  etc. However you can NOT calculate the proportions of genotypes from the allele frequencies (or vice versa) using the H-W law. You have to use the "seat-of-the-pants" method as described below.

**d. What good is all this?** Can use H-W law two ways.

(1). To check if population is in equilibrium.

(2). If you know population is in equilibrium, can use H-W law to figure out proportions of carriers, affected people etc.

Important: This whole business is not as trivial as it looks (if you know algebra) and not as intimidating as it looks (if you are not happy with math). This is illustrated in examples below.

## C. Implications of H-W Law for Evolution

Assuming random mating, no selection and so on, frequency of A and a do not change over time -- A doesn't spread and a doesn't get lost. Even though a heterozygote has the A phenotype (If A dominant to a), allele a is passed on as well as allele A. So variation doesn't get lost -- it is not mixed in and diluted out as in blending inheritance. This is H & W (& Mendel's) big insight and is critical to evolutionary arguments. Darwin didn't know about particulate inheritance and it caused him a lot of worry -- he didn't know why variation stayed around. (Mendel knew about Darwin but not vice versa.)

## D. Some examples of use of H-W

### 1. How to use frequencies, assuming equilibrium

Cystic Fibrosis = CF = most common genetic disease among whites. It's recessive. Person with disease is aa.

Know CF = 1/2000 live births. How many carriers (heterozygotes without symptoms) are there?

$f(a)^2 = f(aa) = 1/2000 = 5 \times 10^{-4}$ ; therefore  $f(a) = \sqrt{5 \times 10^{-4}} = \text{about } 2 \times 10^{-2}$ .

$2pq = 2 \times (1 - 2 \times 10^{-2}) \times 2 \times 10^{-2} = \text{about } 4 \times 10^{-2} = 4/100 = 1/25$ . That's a lot of people.

Over 300 million people in USA, so at least  $10^7$  are carriers; therefore screening is worth it.

At the moment screening is not yet feasible in general population because several different mutations can cause CF. It's easy to screen for those who have the commonest mutation, but that picks up only about 75% of the carriers. It is relatively easy to screen for carriers in families who already have one affected individual, because in these cases the mutation involved is known.

For a similar problem on how to use the H-W see 14-1. For more practice with the H-W law, try 14-2, 14-3, and 14-6 parts A. & B.

2. How to use the H-W law to test for equilibrium. 2 examples with MN. (M & N are surface proteins found on blood cells. M and N are coded for by two co-dominant alleles of the same gene. See problem 9-2.)

a. **The method.** You don't need to check each generation separately to see if we have reached equilibrium. Only need to check proportions of genotypes and alleles of the total population and see if it all fits the H-W distribution.

b. Here are two cases:

**Case 1** (real situation in US)

MM = 30%

MN = 50 %

NN = 20 %

$f(M) = \sqrt{.3} = .55 = p$

$f(N) = \sqrt{.2} = .45 = q$

$p + q = 1$ , ok.

This population is in genetic equilibrium.

**Case 2** (fake)

MM = 25 %

MN = 60 %

NN = 15 %

if plug in,  $f(M) = \sqrt{.25} = .5$

$f(N) = \sqrt{.15} = .4$

$p + q$  not = 1, not ok.

This population is not in genetic equilibrium.

c. **Is non-equilibrium possible?**

The second example (case 2) is given to show that not every population is in genetic equilibrium. When you first see this set up, there seem to be so many variables that it seems impossible not to be able to get everything to fit. But case 2 illustrates that you may have proportions of genotypes that are not compatible with equilibrium. In other words, it is NOT possible (in case 2) to find values of p and q such that  $p + q = 1$  and  $f(MM) = p^2$  etc.

d. **How do you get p and q if there is no equilibrium?**

In case 2 have to use the bean counter or a "seat of the pants" method to get the actual frequencies of M and N. You can NOT use the H-W law to get p and q.

How to do it: Figure out how many slips in the "pot" have M on them and how many have N. One way to do this is to start with 100 people, and assume each person contributes 2 slips to the pot for a total of 200. If person is MM, both slips have M on them, if person is MN, half the slips have M on them, etc. In case 2, number of slips with M = 50 from MM's, and 60 from MN. So  $f(M) = (50 + 60)/200$  etc. Note that in case 2, the values of p and q are the same as in case 1. However the population structure (or genotype of the population) is NOT the same -- the genotypes are different even though the allele frequencies are the same. For some similar examples, see Sadava fig. 21.6 (22.6)

For problems on how to use the H-W law to test for equilibrium, try 14-4 (A & B), and 14-14 part B.

## V. How is all this related to evolution?

A. **Intro: If a population is not in equilibrium, as in case 2 of MN, why not?**

We must examine the assumptions underlying H.W. We said, if there is no selection, etc., then A and a (or M and N) will stay at the same frequency forever and proportions of AA, Aa (or MM, MN) etc. will be constant too (and we know what values they should be). If frequency A and a is changing, or proportions AA (or MN) are wrong, then some of the assumptions must be wrong, and we may have evolution, not equilibrium (or we have non HW equilibrium).

B. **Consider the assumptions behind H-W**

- *Random mating*
- *No mutation or migration*
- *No selection*
- *Large population*

C. **Are all 4 of the assumptions ever true?**

Probably not! But for any particular gene, changes due to one factor (say migration or mutation) may be balanced by changes in the opposite direction due to another factor (say selection). So frequency of alleles of that gene may reach an equilibrium value and stay there (as has happened for genes involved in CF, PKU, ABO, etc.) In other words, genetic equilibrium (like chemical equilibrium) may represent a balance between opposing forces, not a static situation.

D. **If there is no genetic equilibrium, why not? What are the possible causes? Which of the assumptions listed above don't apply? Could be:**

- *Nonrandom mating*
- *Small population*

- Selection
- Mutation/migration

How each of these factors affects the distribution of alleles and genotypes is discussed in detail below.

### E. What is Evolution?

1. *One (reductionist) definition is:* Evolution = change in genotype of a population. Meaning a change in the proportions of A and a in the population. How this happens is discussed below.

2. *Evolution is often divided into two aspects:* Micro vs macro evolution. What's the difference?

a. **Macroevolution:** How do you get from 1 species → 2? Dinosaurs to birds? Are all living things descended from one common ancestor?

b. **Microevolution:** How do you get from light colored bears → dark colored bears? Sensitive bacteria to drug resistant bacteria? Why does one allele (say i) get more common while another (such as I<sup>B</sup>) gets rarer?

3. *Current approach:* We will stick mostly to microevolution, but macroevolution is thought to work in a very similar way, just over longer time periods and involving greater changes in the DNA.

Note: Next term we will work in a little about how molecular structures of eukaryotic cells imply macroevolution and common descent of all living organisms.

## VI. Consider the possibilities (for causing deviations from H.W. equil.) and their consequences, 1 at a time.

**Significance of NRM, mutation & chance (A-C) are listed below. Additional details for A-C are included FYI; we will focus on selection (D).**

### A. Nonrandom Mating (NRM)

1. *Definition & NRM = not mixing the pot.* The probability that A will meet a to → an Aa person is not simply proportional to the frequencies of A and a in the pot.

2. *An example:*

Suppose that MM mates with NN preferentially ("opposites attract"). Or conversely, suppose MM prefer to marry MM and NN prefer NN etc. ("birds of a feather flock together"). Either case is nonrandom mating, and proportions of MM, NN etc. will not be as predicted by H-W, but freq. of M and N in the pot will not change (assuming no selection = everyone is equally likely to marry, and # kids/marriage constant -- see below). This means we don't have evolution -- no change in make up of gene pool or pot. We have a kind of equilibrium, but it is under non standard conditions and it is not what is usually called "genetic equilibrium." It is called non-HW equilibrium or an unusual population structure. The population is in equilibrium, since frequencies of alleles and genotypes don't change, but it's "non H-W" because the proportions of genotypes don't fit the H-W law.

Note that the first possibility discussed above, "opposites attract" is one possible explanation for case 2 (with MN) above.

3. *How does NRM compare to selection?*

NRM = selection of mate is not random with respect to genotype (= unequal mix of pot); however all individuals are just as likely to find a mate and pass on their alleles. There is no favoritism in who gets to reproduce. The only favoritism regards whom you chose for a partner.

Selection = nonrandom selection of alleles to be passed on (= favoring pick of A over a or vice versa either directly or otherwise). In this case, there is favoritism in who gets to reproduce. Ex. of selection as typically stated -- MM have more kids on average than NN. (This case changes what's in the pot in the next generation. NRM does not.)

4. *Significance of NRM.* Nonrandom mating alone doesn't change the proportions of, say, M and N in the pot, but it does affect the chance that M will meet another M (or N) and thus affects the relative proportions of MM, NN and MN.

### B. Mutation & migration

1. *Mutation = Changing slips in pot;* this is very important in the long run but doesn't drive the system in the short run. Amount of current variety is enormous. (This shown by electrophoresis of proteins, restrict. fragment analysis of DNA, etc. Something like 1/100 bases of people are different.) Mutation has been going on so long that the current supply of variation is enormous, and the variety added by ongoing mutations is small. Most 'new' variants are not new mutations but new combinations of old mutations generated by sexual reproduction. So in most cases, changes in a population are likely to be due to changes in the proportions of mutations that are already present, not to new mutations.

Another way to ask this: Does change need to wait for "right" mutation or is the variant usually already there? Maybe, in some cases, esp. with haploids, you do need to wait for mutation to occur. But generally thought not. (Because of high variation levels in all existing populations.)

2. *Migration (also called gene flow) = adding to one pot (= one interbreeding population) from another one; effects are important*

but relatively obvious.

3. **Significance of mutation:** Mutation is the ultimate source of all variation, but the big question is, why do some mutations spread while some die out?

### C. Small populations and/or sampling errors =

1. **Basic idea** -- small samples (of people, gametes, coins, etc.) are usually not exactly representative of the total pool. Ex: flipping coins. If flip only a few, don't get exactly 1/2 and 1/2 heads and tails. If Aa person has a few kids, won't pass on exactly 1/2 A and 1/2 a.

2. **Definition.** Statistical fluctuations in allele frequency that occur as a result of only picking a few alleles from the pot to pass on from generation to next = genetic drift. This can be a one time sampling error (as in Founder effect or the Bottleneck effect) or the cumulative effect of repeated sampling errors that just happen to go in the same direction for several generations. (See below for details.)

#### 3. Types of Drift.

a. **Founder effect** -- what happens when a small group goes off to start a new colony or population. Small group of founders may not be a representative sample, just by chance. One way to state f.e. = starting population usually has less variation than population it came from. Founder effect means population starts with "atypical" gene pool, relative to where it came from. For an example, see Sadava fig. 21.10 (22.10)

#### b. Bottlenecks

Bottlenecks = severe reductions in size of population. Bottlenecks can change the allele frequencies in two ways:

(1). **Bottleneck effect.** See Sadava figs. 21.8 & 21.9. (22.8 & 22.9) What gets through the bottleneck is atypical, so the population "starts over" with a gene pool that is different from the original population. This is similar to the f.e. as described above. (In the case of either the founder effect or a bottleneck, you "start over" with a small atypical population derived from a bigger one. . In the case of the f.e., the small population leaves the bigger one and starts a new group. In the case of a bottleneck, the population stays in place but is greatly reduced in size by plague, predation, war, etc.)

(2). **Statistical Fluctuations.** The population remaining after the bottleneck may be so small that statistical fluctuations in the future may cause significant changes in allele frequencies. (This is genetic drift in the "specific sense" as defined below.) Usually small fluctuations in allele frequency from generation to generation cancel out, but sometimes they are (by chance) all in the same direction, and therefore add up.

c. **Terminology:** Genetic drift can be used two ways, in a general or specific sense.

(1). **General Sense:** Genetic drift can be taken to mean any kind of change in allele frequencies due to small sample size. This is the usual textbook definition, and includes all the changes in small populations due to bottlenecks, founder effects, and statistical fluctuations.

(2). **Specific sense:** Genetic drift can mean the change in allele frequencies due to the cumulative effect of statistical fluctuations from many generations of small samples (which all just happen by chance to deviate in the same direction).

The second type of genetic drift occurs *after* a population passes through a bottleneck (for the usual reasons -- plague, war, or because of founding of a new small population). It might be better to call the second kind of genetic drift "a multiple generation drift effect" or some other specific term. It can be layered on top of a founder/bottleneck effect (see above) or it can occur independently. For example, if a small population (of founders or survivors) is typical, there is no founder or direct bottleneck effect, but allele frequency can drift up or down due to statistical fluctuations as long as the population remains small.

4. **Significance of genetic drift.** In a small population, changes in allele frequencies (either up or down) can occur due to chance alone (no selection involved).

**If you are interested, see problems 14-5 (A & B) & 14-6 (esp. C) for how to tell these various kinds of genetic drift apart.**

### D. Selection -- why this is clearly a major player (or chance vs selection)

1. **How important is chance?** Genetic drift in small populations does occur, and other effects of chance are important too (asteroids?). Is chance a major factor in change that has occurred? Or from another angle, is most of variation you can see now from species to species or individual to individual due to chance or to selection? Exact proportion is debatable, but even if many differences are due to chance, a major proportion of the differences are attributed to selection.

#### 2. Why invoke selection?

a. **Most phenotypic differences seem adaptive** (= make sense, don't seem random) -- match differences in environment, life style, etc. Variants are not randomly distributed. Examples

(1). Bears -- light colored ones with heavy fur are in icy North -- Alaska; darker ones with not-so-heavy fur are in temperate forests -- Yellowstone. Selection explains this; chance doesn't.

(2). Variations in introns vs. exons. There is more variation in intron sequences and in the 3rd position of codons than there is in exons overall. Why should this be? Because (conservative) selection acts against most changes in exons that alter amino acid sequence while there is little or no selection favoring (or disfavoring) changes in introns, spacers, etc.

**b. Selection works well in lab/life** -- breeding of domestic plants & animals, DDT resist. flies, color of moths etc. Works with domestic dogs, pigeons, cows, corn, etc. (intentional) or drug/pesticide resistance (unintentional). Here selection favors new properties, as vs. retention of the old, and is sometimes referred to as innovative selection. For examples, see Sadava figs. 21.2, 21.4, & 21.5 (22.2, 22.4 & 22.5).

### 3. How does selection work?

**a. Basic scheme (long version)** that fits all the data is as follows:

**(1). Random variation occurs by mutation; sex reshuffles the variants.** Mutation is random with respect to function; the environment does not cause the variation to suit need. (Mutation = changes in DNA sequence or number of copies due to mistakes in replication, repair, crossing over or distribution.)

**(2). Selection acts on phenotype (not genotype or individual alleles).** Selection is NOT random with respect to function. Selection = differential reproduction. Different individuals make unequal contributions to gene pool of next generation (some genes/alleles passed on more, some less); no change in genes or individuals.

- Changes that improve chances of survival & reproduction are more likely to be passed on. (innovative selection). That's why bacterial drug resistance spreads. (Rare variants that work better than average become common.)
- Detrimental changes are not likely to be passed on, so working proteins, structures etc. are maintained (conservative selection). That's why cytochrome C stays the same. (Variants that don't work well are lost.)
- Changes in genotype that have no effect on phenotype (neutral mutations) are not selected for or against. These changes tend to be passed on, and it's chance whether they become more or less common. That's why there is more variation in introns than in exons & many synonymous changes in codons.

Note: If A is dominant to a, both AA and Aa will be equally "fit" and equally likely to pass on their alleles. (This is why "a" doesn't disappear.) This is because fitness (and therefore ability to pass on alleles) is determined by the phenotype of the whole organism; not by the genotype or by individual alleles.

**(3). Genotype of population changes** -- Individuals don't change, but the makeup of the gene pool in the next generation is different.

**(4). Adaptation occurs.**

- individuals don't adapt; populations do.
- Fit of population to environment improves.

**b. The short version -- an alternative way to write the basic scheme:** random mutation → random changes in DNA (genotype) → random changes in phenotype → selective reproduction (not random who has the most offspring) → change in genotype (& phenotype) of population due to conditions of life = adaptation

**c. An example:** Up North, light colored (& heavy furred) bears are more likely to have kids and pass on their (light) color genes (& heavy fur genes) so light color and heavy fur spread and vice versa in Yellowstone. Bear with dark fur in Artic can't catch any seals -- prey sees him coming and he goes hungry. Bear with thick fur suffers heat stroke in Yellowstone. And so on.

**d. Natural Selection is a Two Step Process.** Mutation provides variation and then the environment "selects" which individuals (based on their phenotypes) will be most likely to pass on their variant genes/alleles. Note this is a two step process -- first variation occurs; then selection acts on the variants in a separate step. The first step is random (with respect to function); the second step is not. The process involves both 'chance' (random mutation) & 'necessity' (nonrandom selection for function).

**For some problems on the role of selection, see 14-9 to 14-12. For problems on selection vs genetic drift, see 14-4 (part C), 14-5, & 14-6 (part C). There are additional problems on population genetics in problem sets 14 & 15 (15-3 to 15-5).**