BST227
Introduction to Statistical Genetics

Lecture 2:

Principles of mendelian inheritance
Mendelian and complex disease models

Housekeeping Details

Homework posted to course website.
Due Monday, Oct 30

Last time

- **Alleles or Variants**: Different forms (i.e. different DNA sequences) of the same gene or genetic locus. Often called A,B or A,a.

- **Polymorphic:** Polymorphic loci have several different alleles. At non-polymorphic loci, there is no variation from person to person.

- **Genetic Marker:** The DNA sequence at a particular polymorphic locus. Markers are inherently categorical.

- **Disease Susceptibility Locus:** A genetic locus thought to be causal for the disease. Markers are always observable, DSL may not be.

Last time

- Chromosomes come in pairs in diploid organisms (e.g. humans)
- **Genotype:** Pair of alleles at a locus (e.g. AA, Aa, aa)
- **Heterozygote:** genotype with different alleles on the two chromosomes (e.g. Aa)
- **Homozygote:** genotype with the same alleles (e.g AA, aa)
- **Phenotype:** An observable characteristic or trait

Topics for Today

Mendel's breeding experiments

How are genotypes inherited?

Models for genotype-phenotype relationship

Biology underlying inheritance:
Crossing-over, recombination

Gregor Mendel

Studied trait inheritance by controlled breeding of ~30,000 pea plants between 1856 - 1863

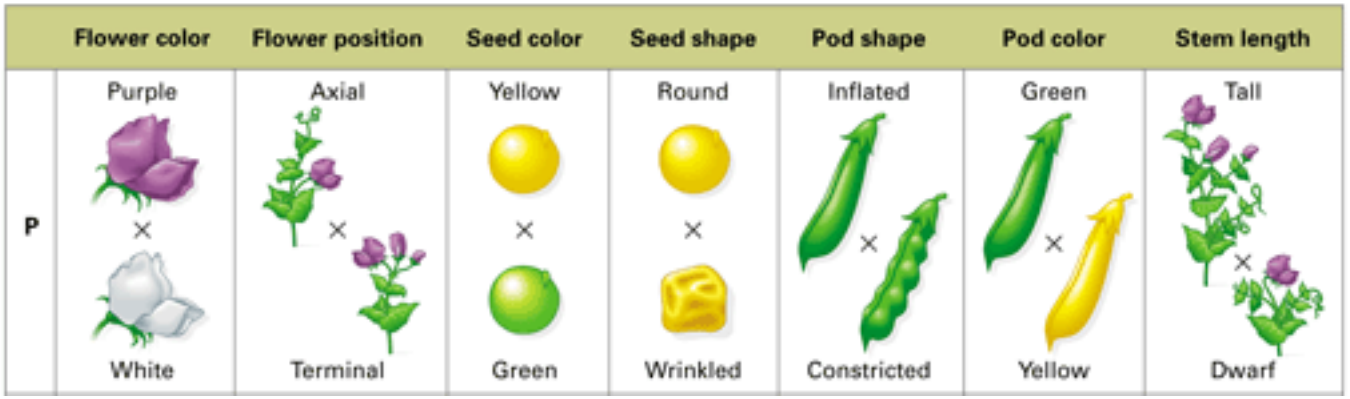
What was known before Mendel's Experiments?

- Existence of inherited traits was known, but not underlying genetic model or mechanism
- Physical units which were the basis of inherited traits (later called genes)—inherited by offspring
- Physical units had different forms (alleles)
- Theory of blending inheritance

Mendel's Methods

Key choices:

1. Dichotomous traits: yellow vs green pea pods, white vs purple flowers, round vs wrinkled seeds.
   "Constant, differentiating traits"

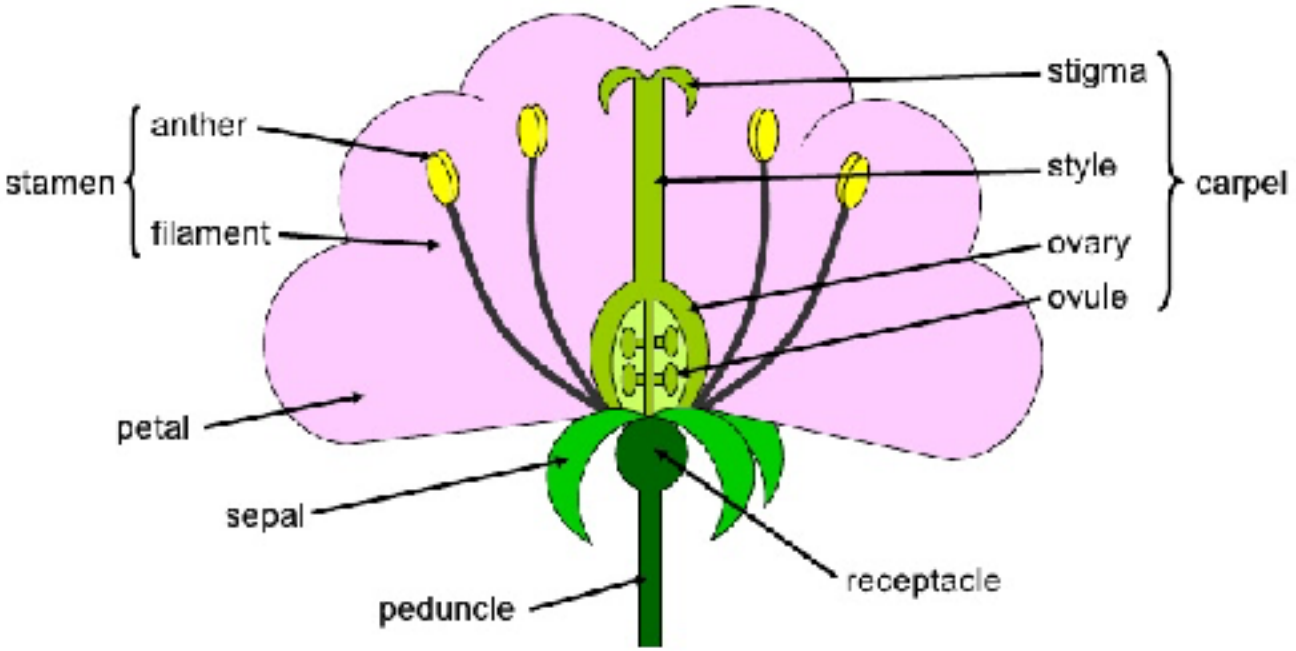


http://guestblog.scientopia.org/wp-content/uploads/sites/35/2012/07/10-04.gif

Mendel's Methods

Key choices:

2. Plants which were both self- and cross-pollinating



http://www.slideshare.net/chduncan/plant-reproduction-12751906

http://guestblog.scientopia.org/wp-content/uploads/sites/35/2012/07/10-04.gif

Mendel's Methods

Key choices:

2. Plants which were both self- and cross-pollinating

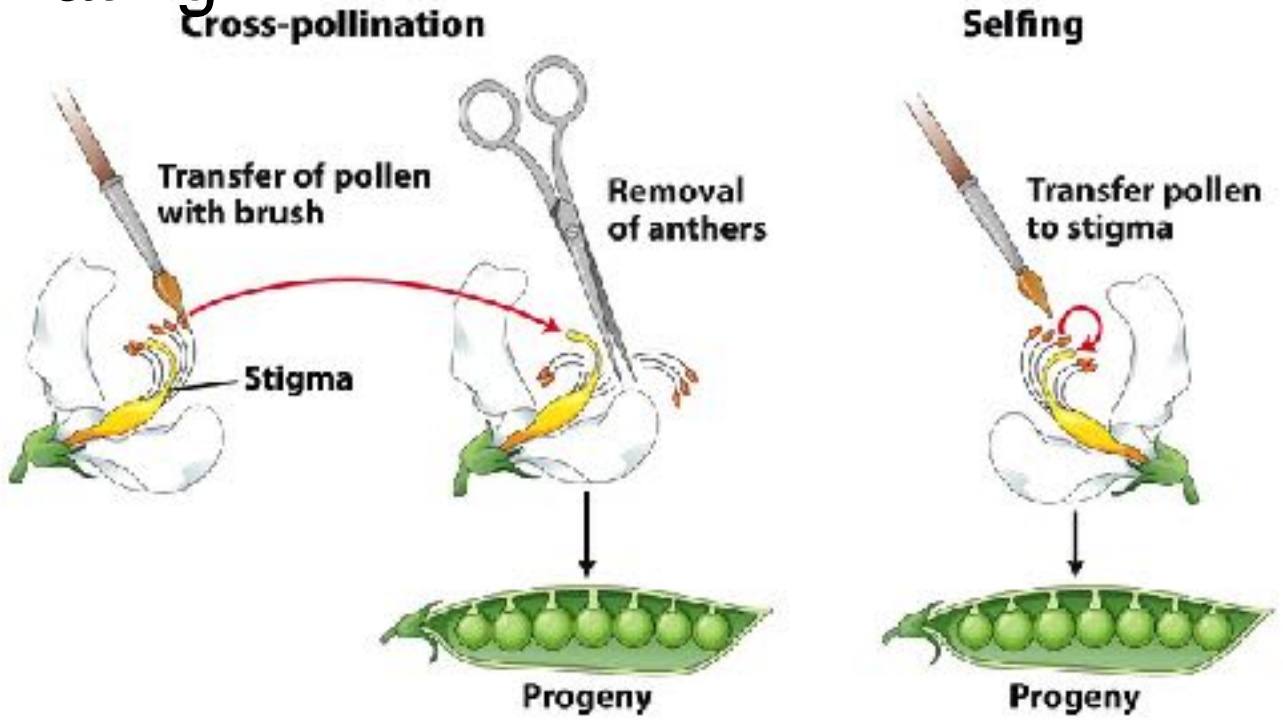
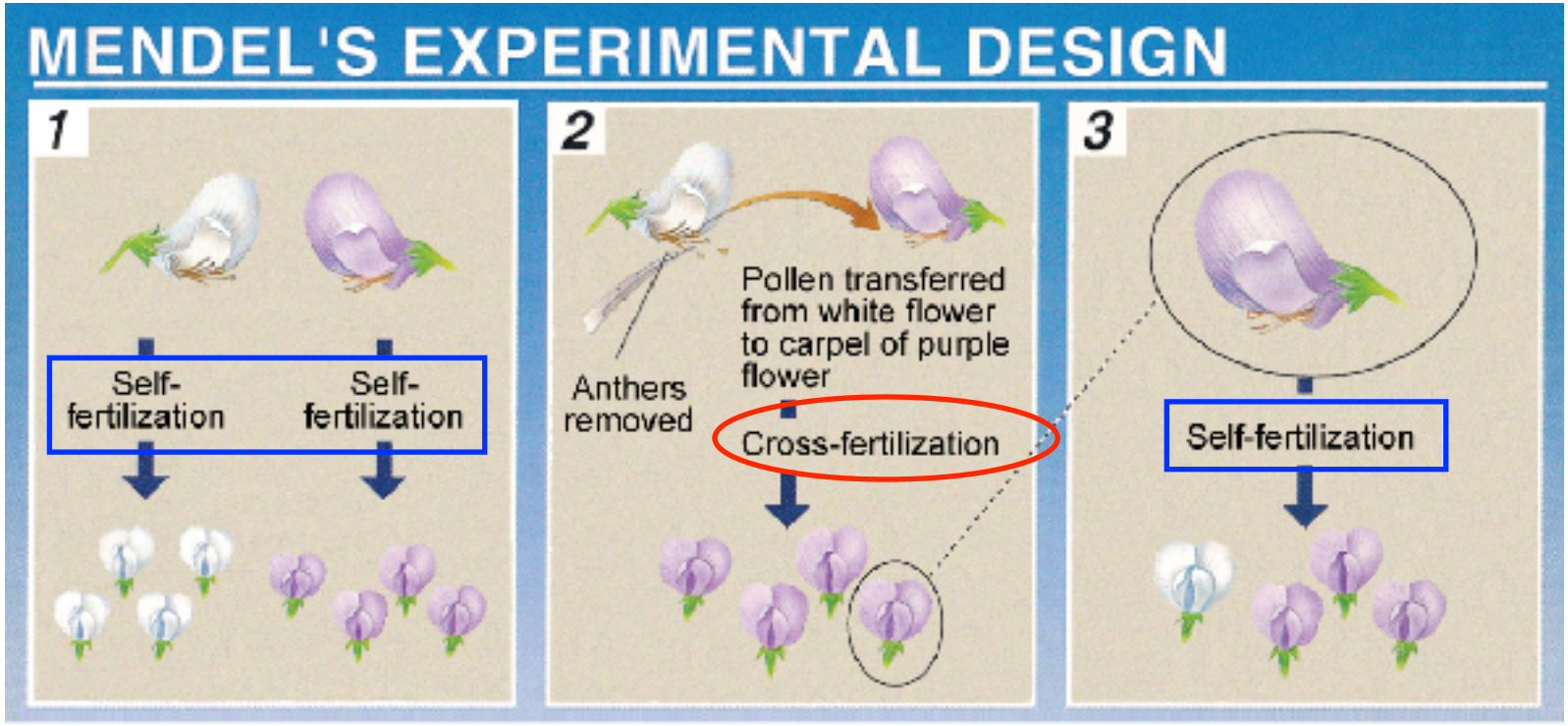
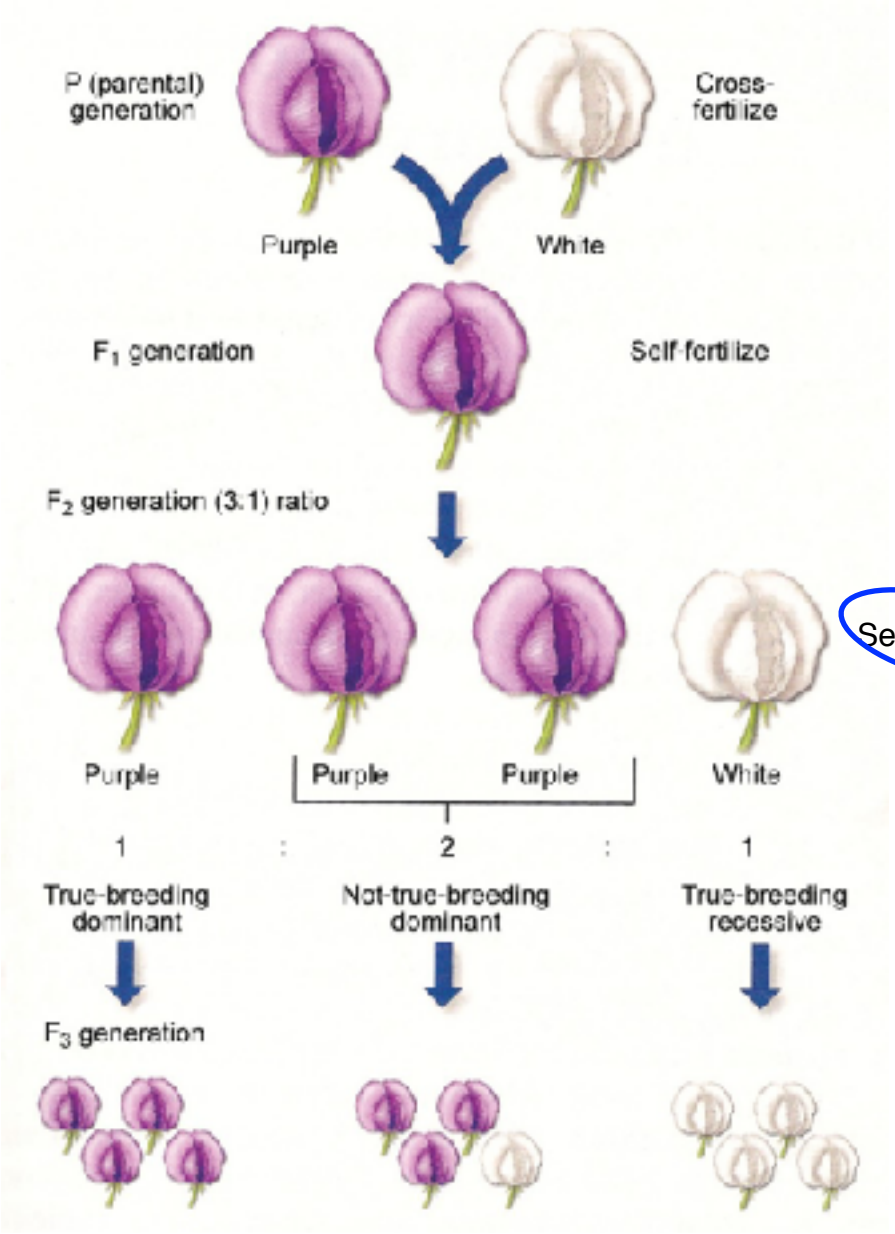


Figure 2.3
Introduction to Genetic Analysis, Tenth Edition
© 2012 W. H. Freeman and Company

http://guestblog.scientopia.org/wp-content/uploads/sites/35/2012/07/10-04.gif





Parental plants are pure-breeding

All F_1 plants have purple flowers

¾ of F_2 plants have purple flowers

Mendel's Models

Assume 2 alleles: A and a

Genotypes are AA, Aa, aa

Models for Phenotypes or traits:

Recessive or Dominant

Assume a dichotomous trait is completely determined by the genotype at a single trait locus

Distribution of offspring's genotype conditional upon parental genotypes

		WHITE	PURPLE	
Father's Genotype	**Mother's Genotype**	**aa**	**Aa**	**AA**
aa	aa	1	0	0
aa	Aa			
aa	AA	0	1	0
Aa	aa			
Aa	Aa	¼	½	¼
Aa	AA			
AA	aa			
AA	Aa			
AA	AA	0	0	1

Distribution of offspring's genotype conditional upon parental genotypes

		WHITE	PURPLE	
Father's Genotype	**Mother's Genotype**	**aa**	**Aa**	**AA**
aa	aa	1	0	0
aa	Aa	½	½	0
aa	AA	0	1	0
Aa	aa	½	½	0
Aa	Aa	¼	½	¼
Aa	AA			
AA	aa			
AA	Aa			
AA	AA	0	0	1

Mendelian Inheritance: Mendel's First Law

Mendel's First Law (Law of Segregation):

Offspring inherit one allele from each parent. Parents transmit alleles independently of each other, and within a parent, transmission is random and with equal probability for each allele.

It follows from Mendel's law that each of an individual's two alleles can be labeled, as their maternal allele or their paternal allele

Mendel's Experiments with Two Traits: Stem length (A,a) and flower color (B,b)

Question: are the genes underlying two traits transmitted independently?

If so, then:

based on independent of A a B b

$$
\begin{aligned}
P(\mathrm{AaBb} \mid \mathrm{AaBb}, \mathrm{AaBb}) &= P(\mathrm{Aa} \mid \mathrm{Aa},\mathrm{Aa}) P(\mathrm{Bb} \mid \mathrm{Bb},\mathrm{Bb}) \\
&= \frac{1}{2} \times \frac{1}{2} \\
&= \frac{1}{4}
\end{aligned}
$$

Mendel's Experiments with Two Traits: Stem length (A,a) and flower color (B,b)

Mendel started with pure forms: tall plants with purple flowers (AA,BB) and short plants with white flowers (aa,bb).

F1 hybrids have one possible genotype and one possible phenotype - What are they?

F2 hybrids have four possible phenotypes in the ratio 9:3:3:1 - What are they?

F2 hybrids have nine possible genotypes in the ratio 1:2:1:2:4:2:1:2:1 (assuming independence) - Try it!

Mendel's Second Law

Mendel's Second Law *(independent assortment)* :

Alleles of different genes are transmitted independently.

Problem: Mendel's Second Law is not always true. Applies exactly when two genes lie on different chromosomes, but can be false if genes are 'close' together on the same chromosome.

Why did Mendel miss? Used 7 traits, and all were NOT on different chromosomes---genes for some traits were not independently transmitted, but Mendel's sample sizes were not big enough to detect it.

Lack of independence is fortuitous—forms the basis of linkage analysis

From Genotype to Phenotype: More about the penetrance function

Definition:

Penetrance function = set of conditional probabilities for the phenotype Y (here assumed dichotomous) given the genotype G,

i.e. P(Y=1|G)

With disease traits, Y=1 usually taken to mean disease.

Assume a gene with 2 alleles:

- d: the common (or wild type or normal or ancestral) allele
- D: the *disease susceptibility allele or disease variant*
- There are three possible genotypes: dd, dD and DD

The Mendelian penetrance functions

- Mendel's penetrance functions were 0/1 (Complete penetrance):

Dominant trait: $P(Y=1 \mid DD \text{ or } Dd) = 1$

D is dominant

$P(Y=1 \mid dd) = 0$

Recessive Trait: $P(Y=1 \mid DD) = 1$

D is recessive

$P(Y=1 \mid Dd \text{ or } dd) = 0$

- Dominant mode: P(disease) requires only one disease allele
- Recessive mode: P(disease) requires two disease alleles
- Examples from last time?

More Realistic Models for Complex diseases

Complex Diseases: Relationship between genotype at a single locus and disease can depend on many other factors.

Phenocopy: P(disease) > 0 in absence of any disease alleles (dd)

Reduced Penetrance: Having the requisite # disease alleles does not guarantee disease

- Dominant: P(Y=1|dD) = P(Y=1|DD) < 1 and P(Y=1|dd) > 0
- Recessive: P(Y=1|dD) = P(Y=1|dd) > 0 and P(Y=1|DD) < 1
- Codominant: Pr(Y=1|dd) ≠ P(Y=1|dD) ≠ P(Y=1|DD)
  all three genotypes have different effects on the disease risk. In most cases:
  P(Y=1|dd) < P(Y=1|dD) < P(Y=1|DD)
  P(disease) increases monotonically with # D alleles
- Additive or Dose Dependent: P(Y=1|dD) midway between Pr(Y=1|dd) and P(Y=1|DD) ("midway" depends on the scale)

(See Homework)

Continuous traits

A common model:

Y is $N(\mu_g, \sigma^2)$: Mean response depends on genotype, g=dd, dD, DD

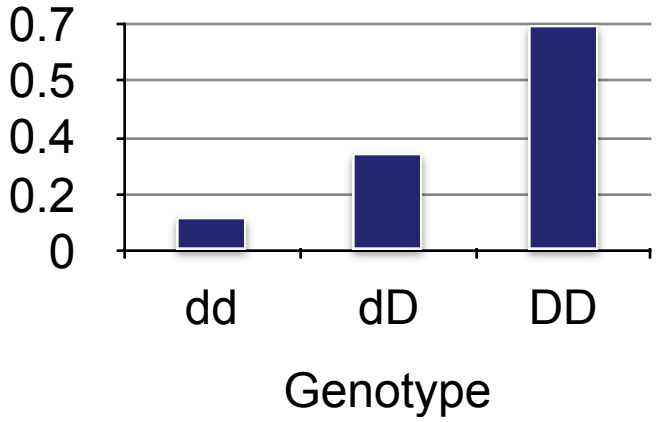
dominant mode of inheritance: $\mu_{dD} = \mu_{DD}$

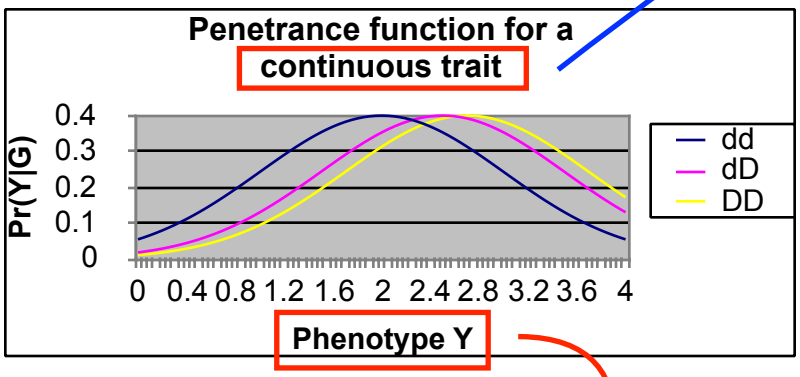
additive: $\mu_{dD} = (\mu_{dd} + \mu_{DD})/2$

Many other possibilities.

Penetrance functions for Complex Traits

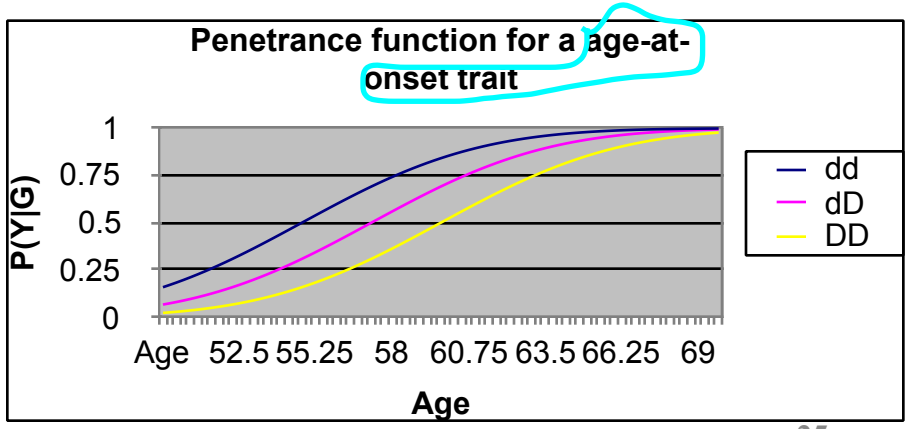
Penetrance function for a binary trait





e.g.blood pressure, BMI

phenotype of continuous measurement like hight, weight



Using Regression Models to Specify Mode of Inheritance

Code genotype as a variable X; coding can be chosen to represent different genetic models;

$$E(Y|X) = \beta_0 + \beta_1 * X$$

Ways of coding X to represent different modes of inheritance ($E(Y|X=0) = \beta_0$):

Recessive		Dominant		Additive	
X	G	X	G	X	G
1	DD	1	DD/Dd	2	DD
0	dd/Dd	0	dd	1	Dd
				0	dd

Co-Dominant (need two dummy variables to code 3 groups)

X1	X2	G
1	0	DD
0	1	Dd
0	0	dd

Generalized Linear Models

(Used to fit models to data and estimate gene effects)

$$g(E(Y|X)) = \beta_0 + \beta_1 * X$$

Where g() is a link function. Linear regression, g is the identity link.

With dichotomous outcomes E(Y|X) = P(Y=1|X), g() often taken to be logistic link function:

$$g(E(Y|X)) = \log[P(Y=1|X)/(1- P(Y=1|X)]$$

Log link can be used for a relative risk model:

$$\log P(Y=1|X) = \beta_0 + \beta_1 * X.$$

Easy to include covariates that might affect the trait, e.g. age, exposure

Biology Underlying Mendel's Laws

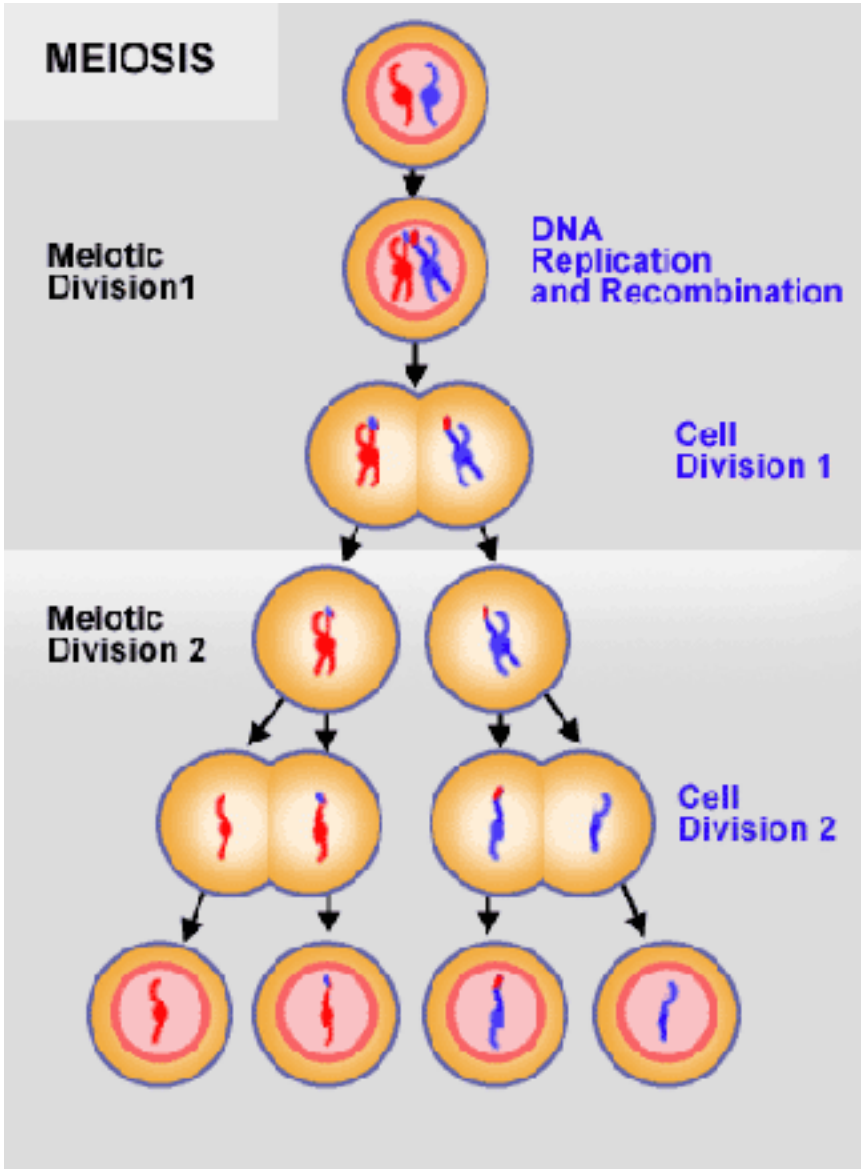
Cell Division

Mitosis: Used for growth and replacement of worn out cells. Creates new cells genetically identical to old ones.

Meiosis: Used for reproduction. Creates gametes (sex cells): either sperm or egg cells. Genetic material in the cells is not identical to parental cells; meiosis generates genetic variability through the process of crossing-over between maternal and paternal chromosomes.

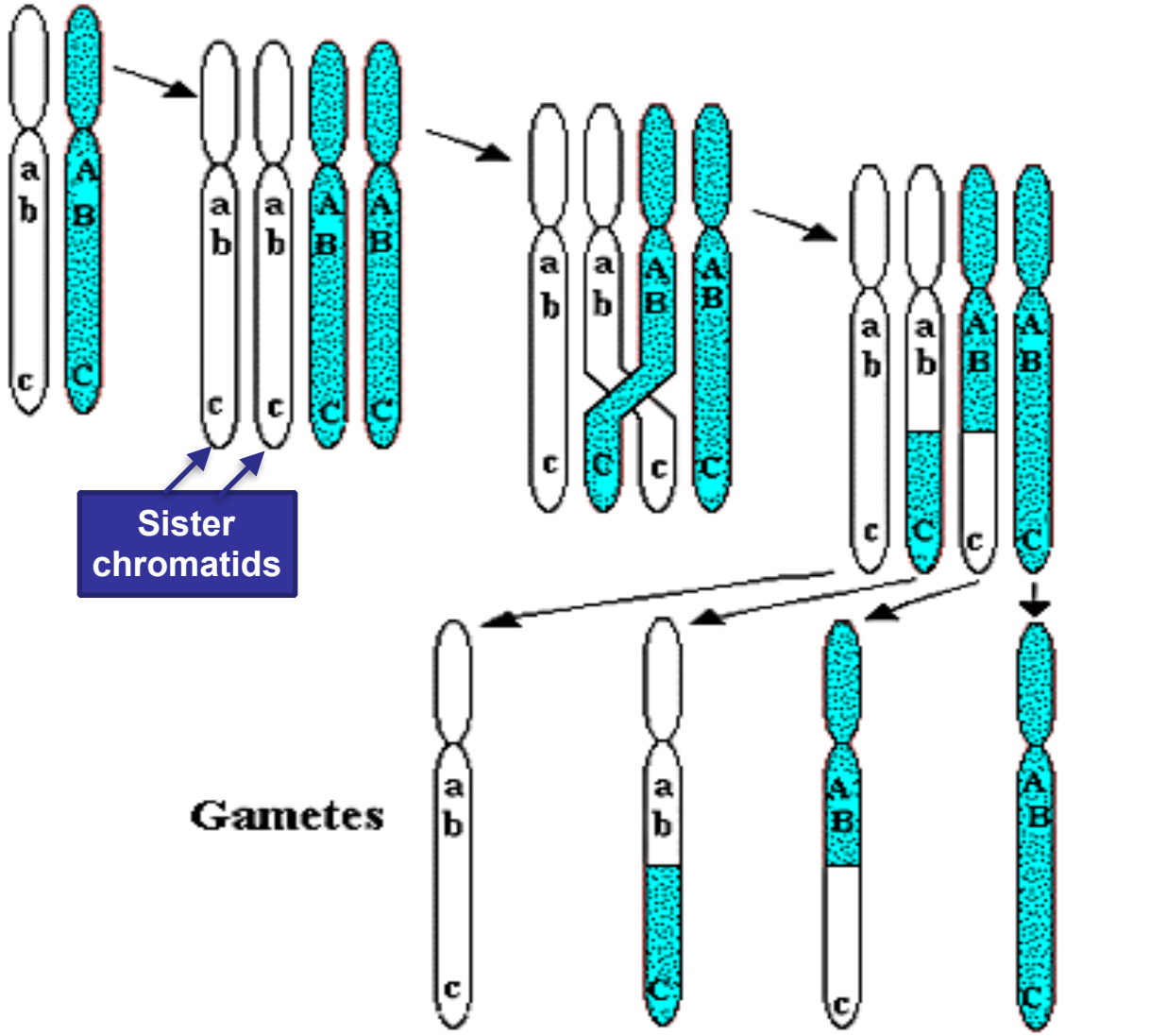
Diploid: cells have **two** copies of each chromosome

Haploid: cells have **only** one copy of each chromosome



http://www.tokresource.org/

Meiosis starts with a diploid cell and makes 4 haploid cells

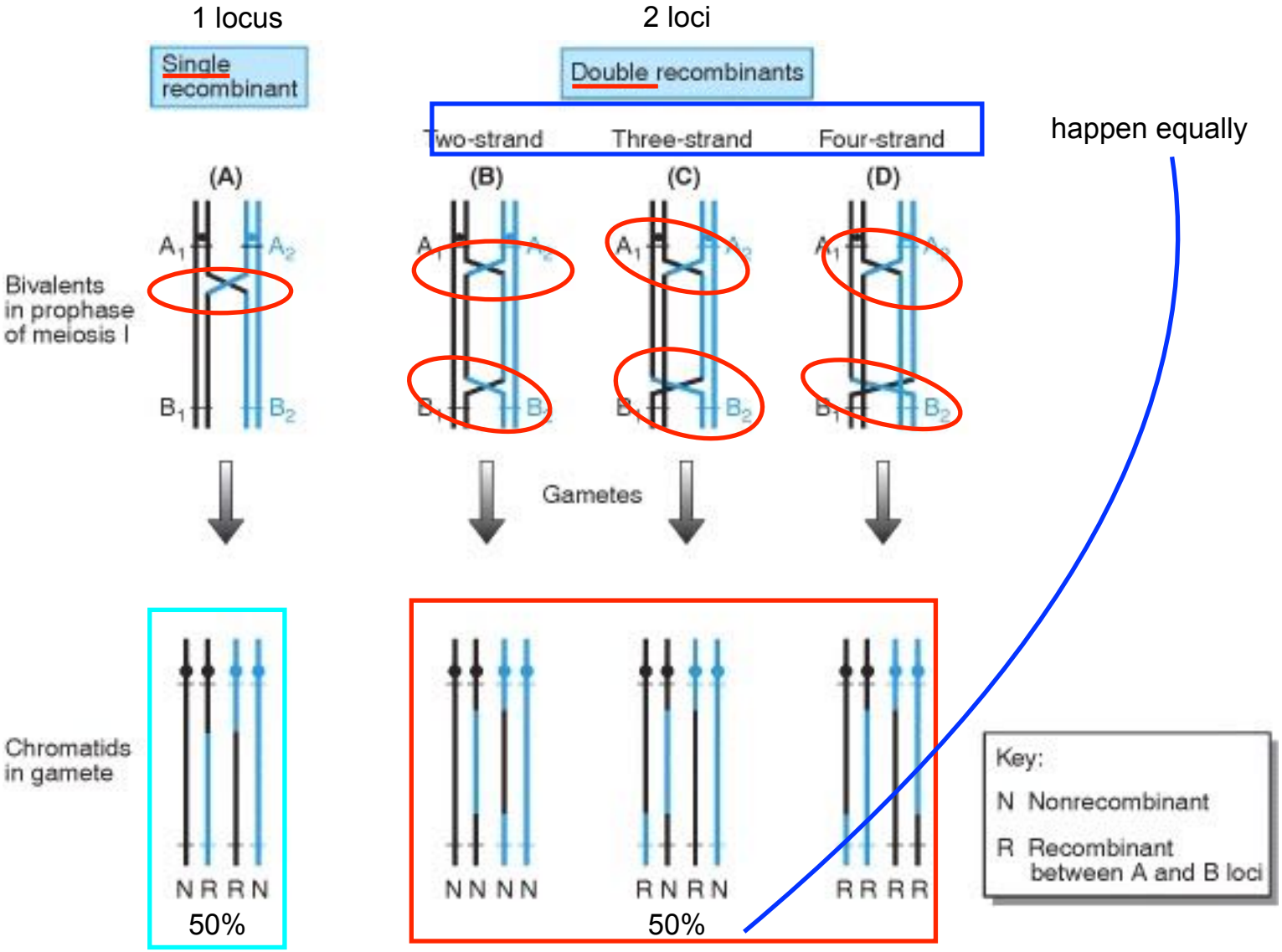


Crossing-over and recombination during meiosis

Adapted from:
Morgan T.H., Sturtevant A.H., Muller H.J., and Bridges C.B., "The Mechanism of Mendelian Heredity", 1915.

Crossing over and Recombination

- Cannot directly observe crossovers. Can (sometimes) observe recombinations
- Recombinant chromosome: Have a combination of alleles in offspring on a chromosome different from what the parent had
- Recombinants and crossovers are not 1-1
- Why?



The average effect of a double crossover is to give 50% recombinants.

Human Molecular Genetics. 2nd edition., Strachan T, Read AP., New York: Wiley-Liss; 1999.
http://www.ncbi.nlm.nih.gov/books/NBK7560/figure/A1383/

Recombination

- Crossovers occur randomly along a chromosome (not independently or uniformly)
- P(recombination event between two loci) is called the recombination fraction (theta).

  P(recombination between 2 loci) = θ
- Roughly speaking, θ measures distance between two loci (on same chromosome)

Recombination

- If two loci are very close together,

  $\theta \approx 0$

  In this case, independent assortment is clearly violated. Why? mendelain law is violeted
- If two loci are very far apart (or on different chromosomes), $\theta \approx ½$ and independent assortment holds. Why does $\theta = ½$ imply independent assortment? 50% recombination

Housekeeping Details

Homework posted to course website.
Due Monday, Oct 30

Next Time

- Population Genetics (chapter 3)
- Concepts of familial aggregation and recurrence risk ratios