AN INTRODUCTION TO RECOMBINATION AND LINKAGE ANALYSIS

Mary Sara McPeek

Presented by: Yue Wang and Zheng Yin

11/25/2002

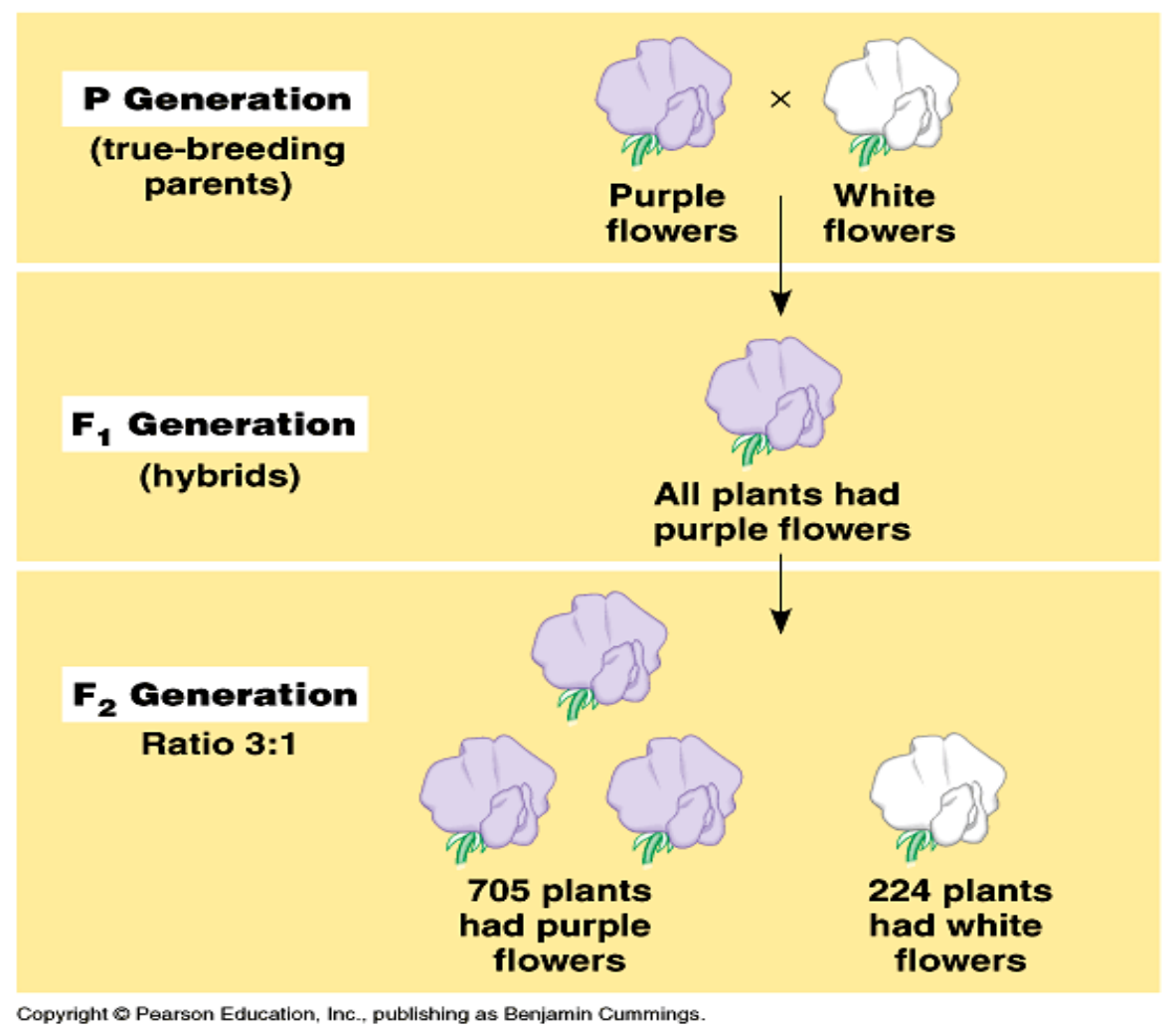
Outline

- Mendel’s Laws
- Linkage and recombination
- Linkage analysis

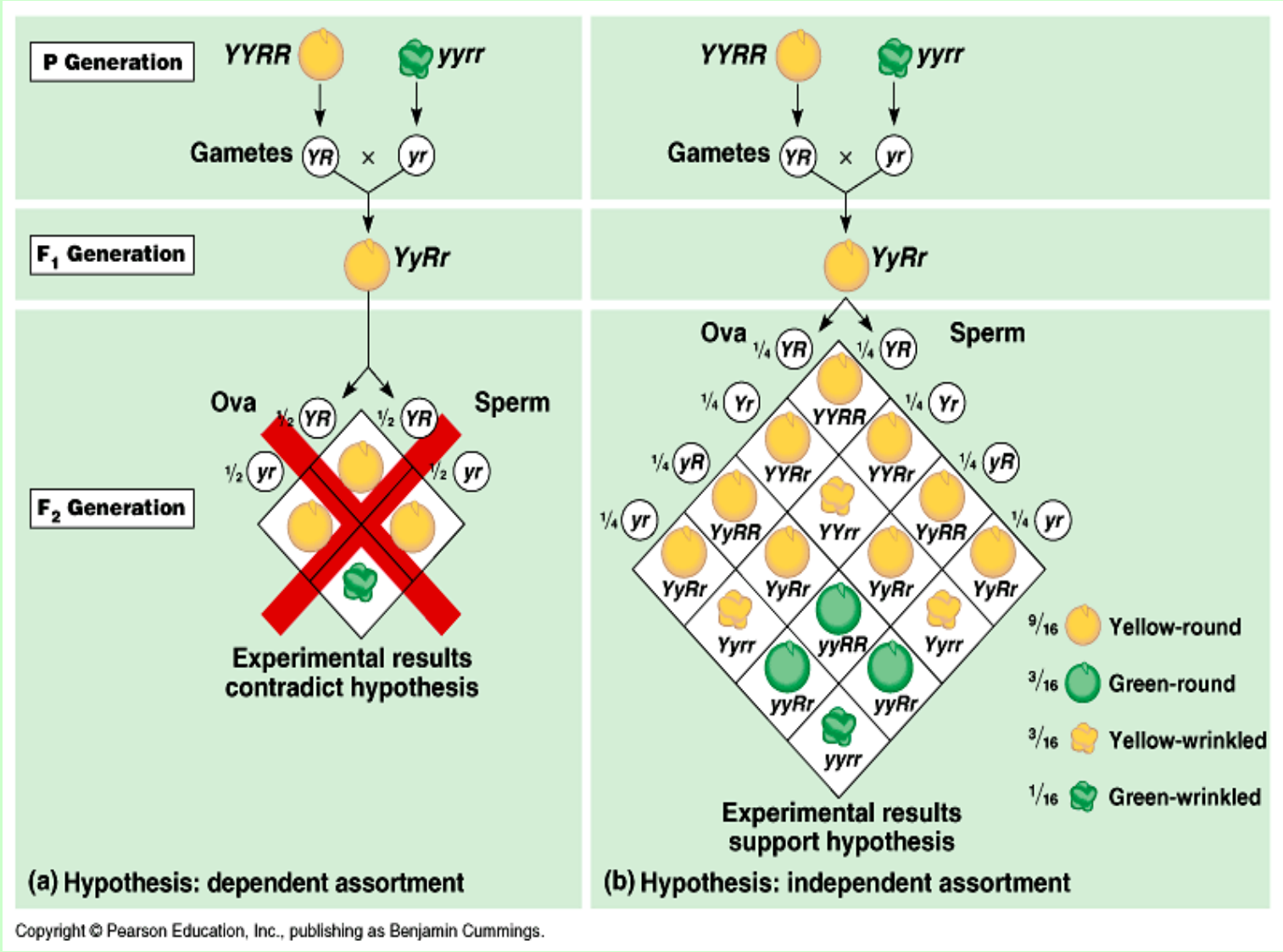
Mendel's Laws

- Mendel's First Law
  - The two alleles for each character segregate during gamete production.
- Mendel's Second Law
  - Each pair of alleles segregates into gametes independently.

Mendel's Experiment 1

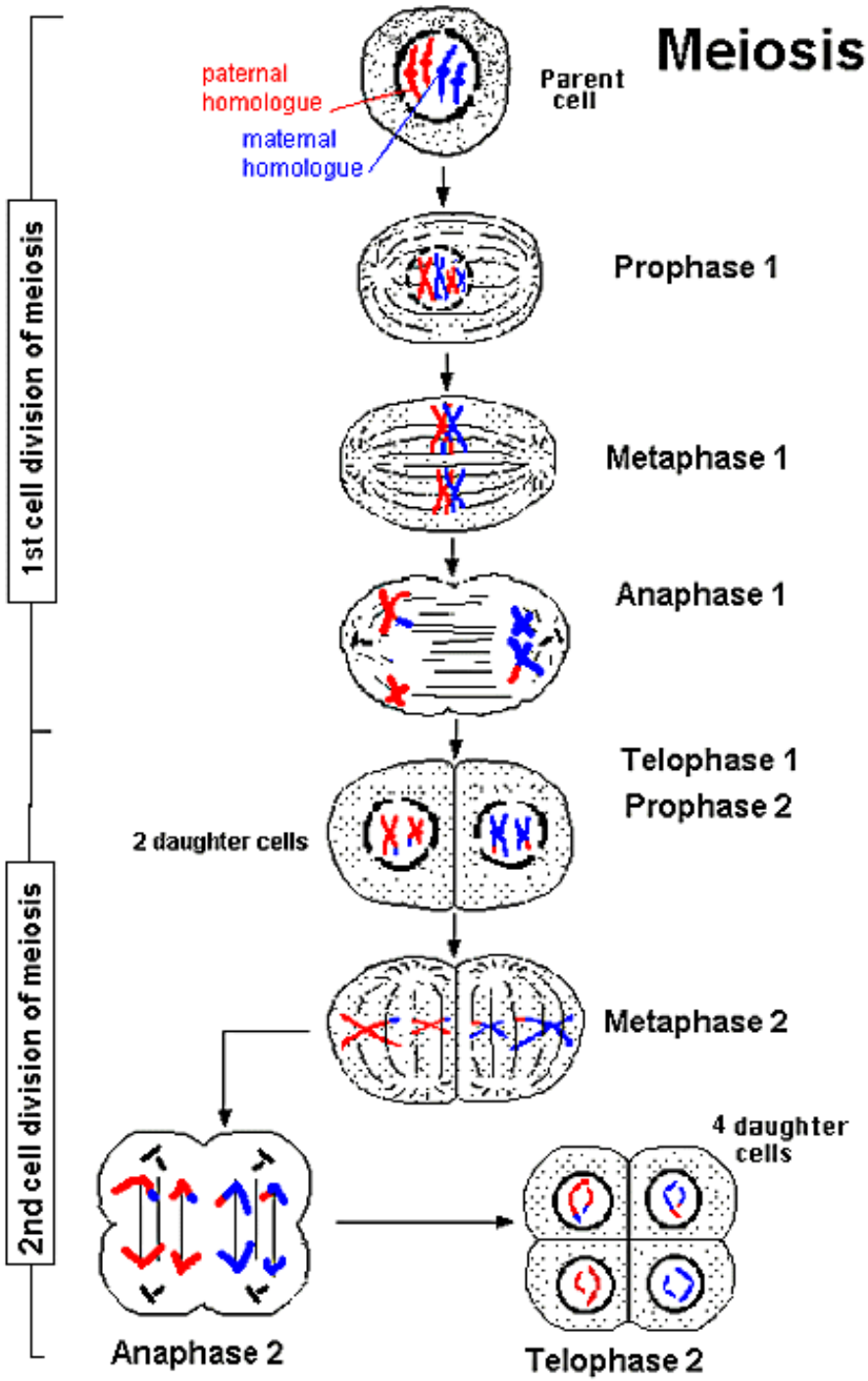


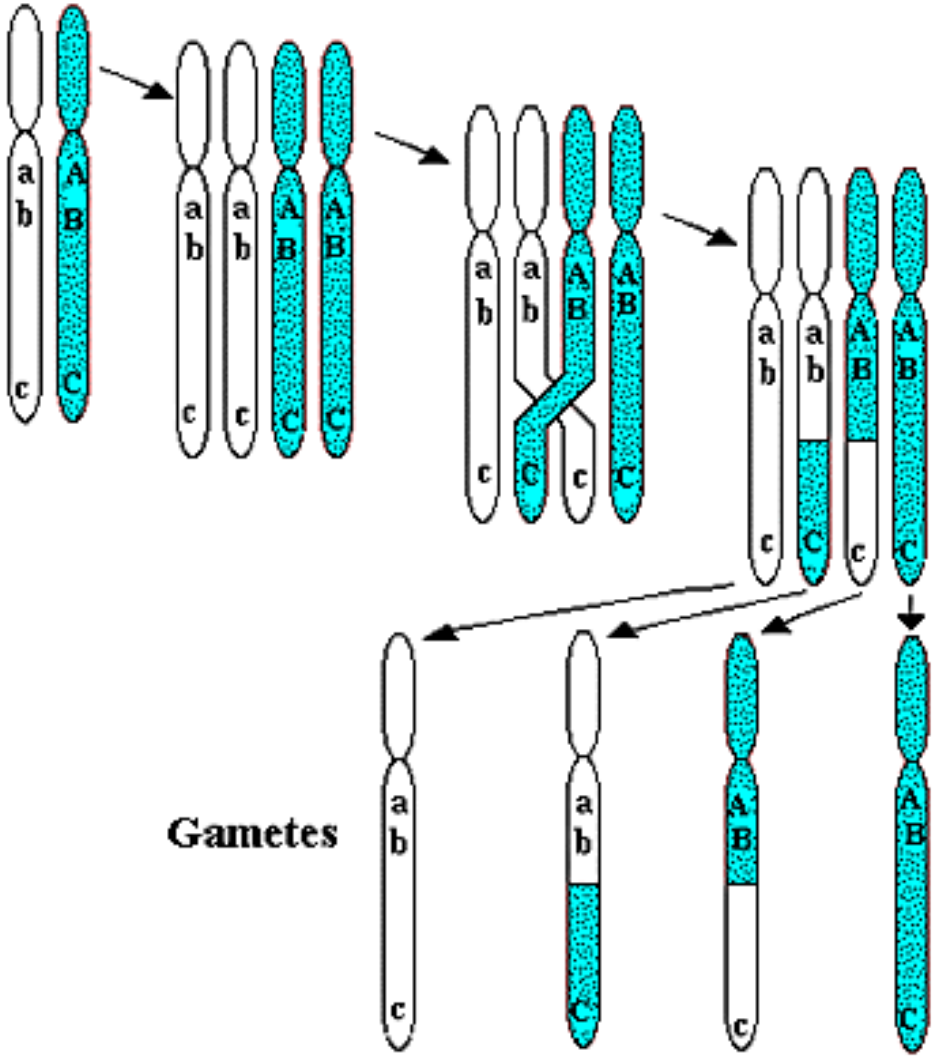
Mendel's Experiment 2



Linkage and Recombination

- Bateson's report and Morgan's Explanation
  - Two characters did not segregate independently, nor were they completely linked.
  - Morgan: Crossover.
- Meiosis process
  - Crossover only occurs during first meiotic division
- Three processes lead to most genetic variation:
  - Independent orientation of chromosomes in meiosis
  - Crossing over of chromosomes in meiosis
  - Random fertilization
- Recombination and Crossover


Meiosis
paternal homologue
maternal homologue
Parent cell
Prophase 1
Metaphase 1
Anaphase 1
Telophase 1
Prophase 2
2 daughter cells
Metaphase 2
4 daughter cells
Anaphase 2
Telophase 2
1st cell division of meiosis
2nd cell division of meiosis



Crossing-over and recombination during meiosis

Linkage Analysis

- What is Linkage Analysis
- Genetic distance
- Map function: $r = M(d)$
- Multilocus Linkage analysis

Linkage Analysis

- Key to linkage analysis:
  - The smaller the amount of recombination observed between genes, i.e. the more tightly linked they are, the closer we could infer that they lie on a chromosome.
- Goal: Place genes and genetic markers along chromosomes, order them, assign genetic map distance.
  - Genetic markers: sequence of DNA with unknown functions but easily recognized as ‘landmarks’

Genetic Distance between two loci

- Definition by Sturtevant (1913):
  The expected number of crossovers per meiosis between the two loci on a single chromatid strand. Unit: Morgan/centiMorgan
  - Known as the genetic map distance: *d*
- Recombination fraction *r*:
  The frequency of meiotic products that are recombinant.

Map Function

- Haldane (1919)

$$r = M(d)$$

$$M(d) = [1- exp(-2d)] / 2$$

  - Relationship between recombination and crossing-over
- No-Interference (NI) model assumptions
  - Chiasma process is a Poisson process: Poisson(2d)
  - No chromatid interference (NCI)
- Mather's Formula (1935):

$$r = [1 – P(N=0)] / 2$$

Derivation of Mather's Formula

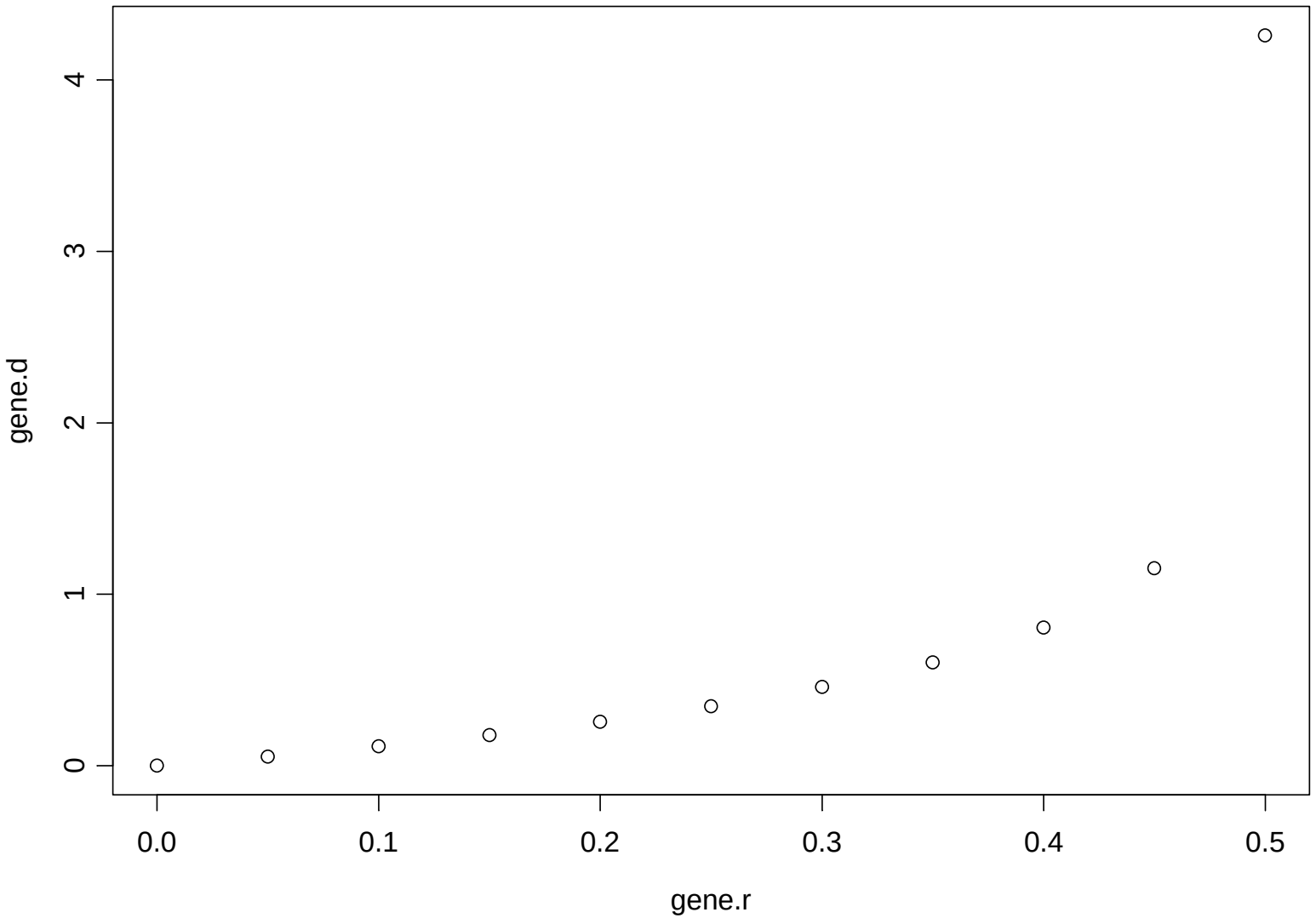
- Under NCI assumption, if there are N crossovers in the chiasma process on an interval, with N>0, the the chance of having i crossovers on a given chromatid is

$$\binom{N}{i} \times \frac{1}{2^i} \times \frac{1}{2^{N-i}}$$

- Recombination will occur in the interval if the chromatid is involved in an odd number of crossovers

$$\frac{1}{2^N} \times \sum_{i=0}^{[[\frac{N-1}{2}]]} \binom{N}{2i+1} = \frac{1}{2}$$

$$d = -ln(1 - 2r)/2$$



Multilocus Linkage Analysis(1)

- Look at a number of loci simultaneously.
  - Assuming that all recombination among the *m* loci could be observed.

    ABC ABC ABc AbC Abc

    a b c a b c a bC aBc aBC

    Data counts: $2^m \rightarrow 2^{m-1}$
- Index the loci by *i*, $i_j = 0$ implies no recombination between loci i_j and i_{j+1}

Multilocus Linkage Analysis(2)

- Fisher (1922): Method of maximum likelihood for linkage analysis.

  Consider *m* loci simultaneously, the NI probability is:

$$p_i = \prod_{j=1}^{m-1} \theta_j^{i_j}(1-\theta_j)^{1-i_j} = \frac{1}{2}\prod_{j=1}^{m-1}(1-e^{-2d_j})^{i_j}(1+e^{-2d_j})^{1-i_j}$$

This probability depends crucially on the presumptive order of the markers.

Multilocus Linkage Analysis(3)

- The likelihood of the data:

$$L(\theta, n) \propto \prod_i p_i^{n_i} = \prod_{j=1}^{m-1} \theta_j^{\sum_{i:i_j=1} n_i} (1-\theta_j)^{\sum_{i':i'_j=0} n_i}$$

- Estimation of the recombination fraction:

$$\hat{\theta}_j = min(\sum_{i:i_j=1} n_i \div \sum_{i'} n_{i'}, \frac{1}{2})$$

Multilocus Linkage Analysis(4)

- Under the NI assumption, reduced to to a pairwise analysis of recombination between adjacent markers.

$$\theta_{AC} = \theta_{AB}(1 - \theta_{BC}) + (1 - \theta_{AB})\theta_{BC}$$

- Estimate order by maximize the appropriate likelihood under each of the candidate orders
- Problems
  - Humans cannot be experimentally crossed
  - Maternal and paternal alleles may be identical at some loci
  - Ancestors may not be available for the analysis

Have a nice Thanksgiving!