Genotype Frequencies

COMP 571 - Fall 2010

Luay Nakhleh, Rice University

Outline

(1) Mendel's model of particulate genetics

(2) Hardy-Weinberg expected genotype frequencies

(3) The fixation index and heterozygosity

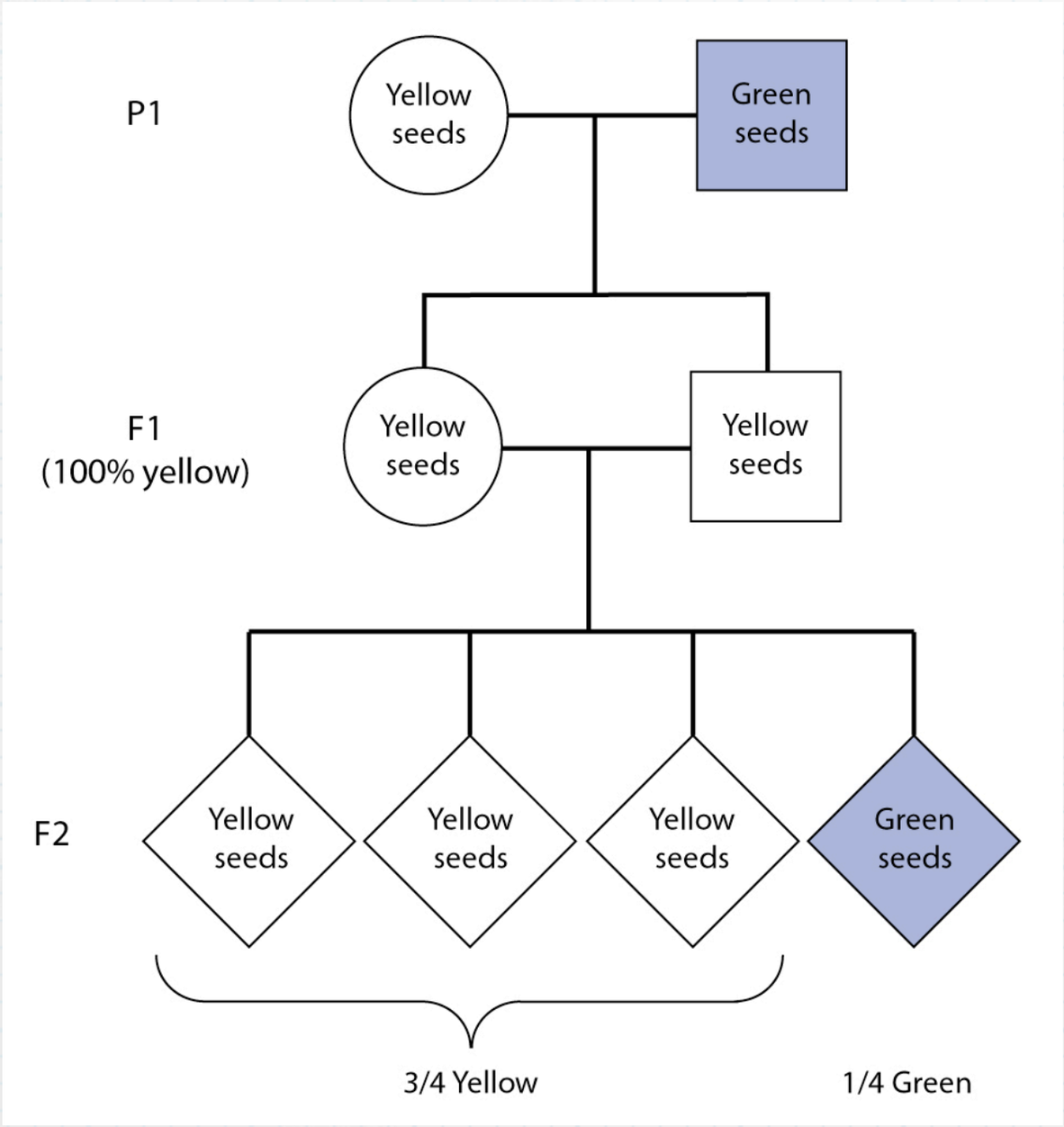
(4) Mating among relatives

(5) Gametic disequilibrium

(1) Mendel's Model of Particulate Genetics

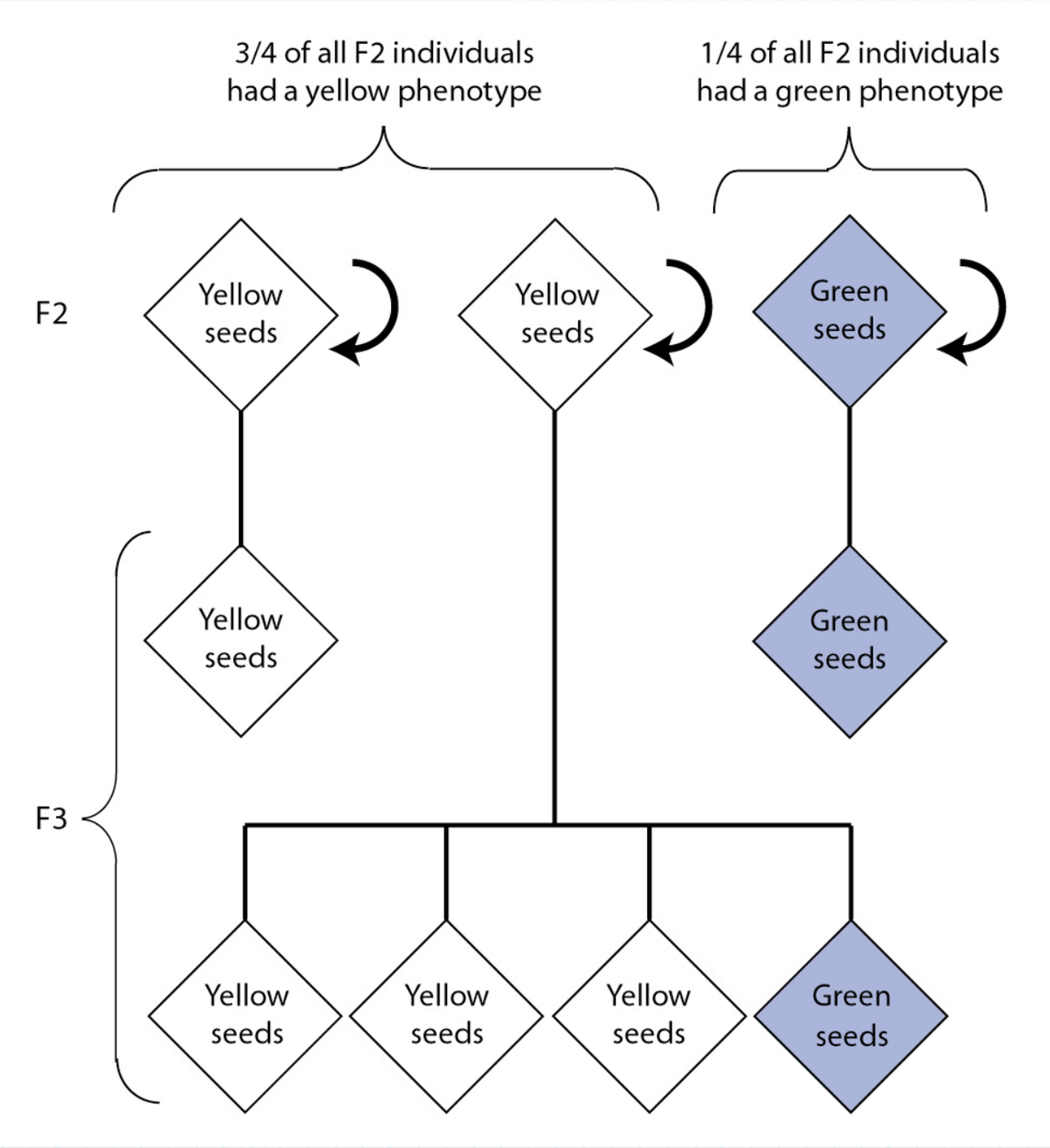
* Mendel used experiments with pea plants to demonstrate independent assortment of both alleles within a locus and of multiple loci
* Mendel used pea seed coat color as a phenotype, and his goal was to determine, if possible, the general rules governing inheritance of this phenotype

(1) Mendel's Model of Particulate Genetics



Mendel's crosses to examine the segregation ratio in the seed coat color of pea plants. The parental plants (P1 generation) were pure breeding, meaning that if self-fertilized all resulting progeny had a phenotype identical to the parent. Some individuals are represented by diamonds since pea plants are hermaphrodites and can act as a mother, a father, or can self-fertilize.

(1) Mendel's Model of Particulate Genetics

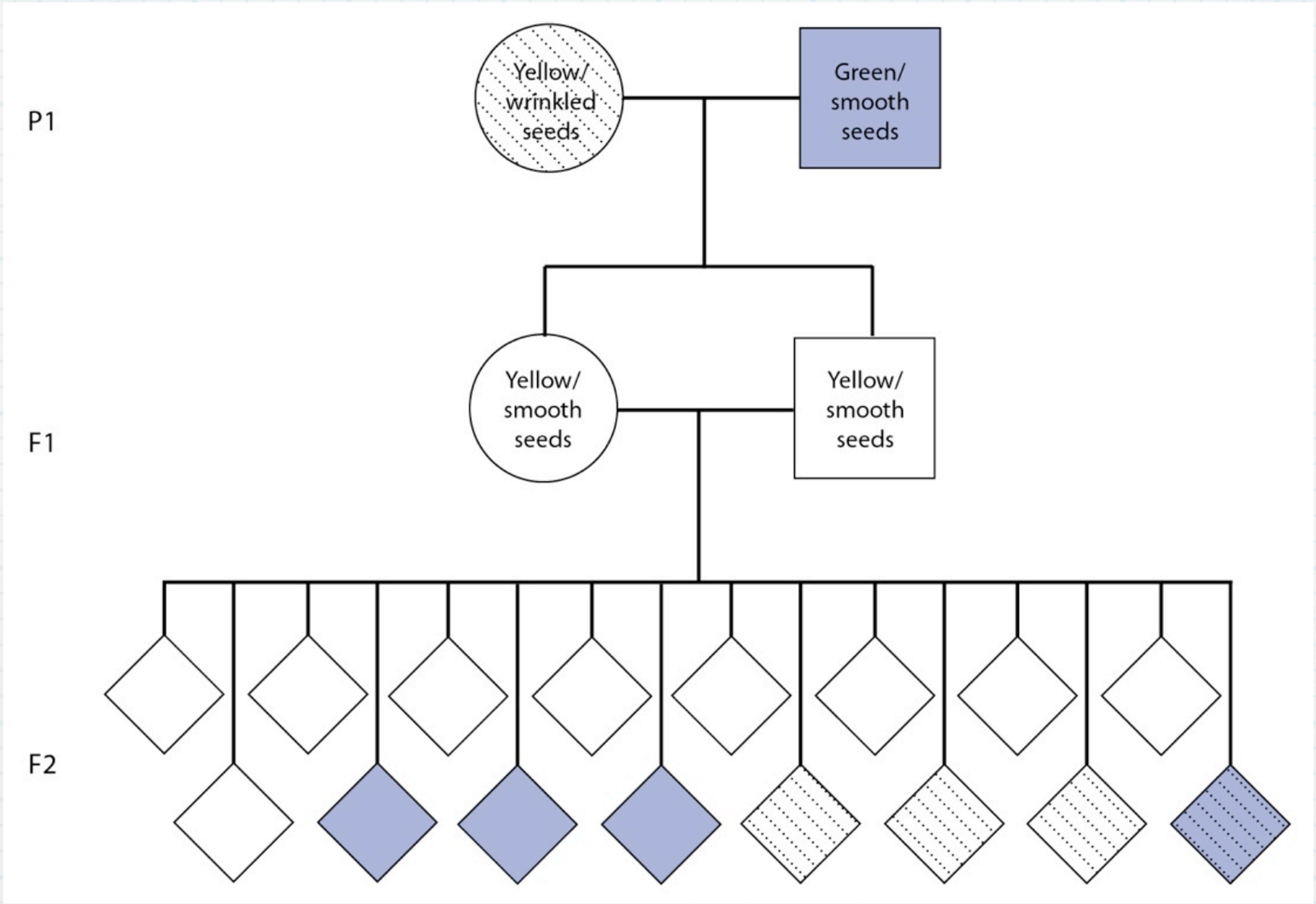


Mendel's self-pollinated (indicated by curved arrows) the F2 progeny produced by the cross shown in the figure on the previous slide. Of the F2 progeny that had a yellow phenotype (three-quarters of the total), one-third produced all progeny with a yellow phenotype and two-thirds produced progeny with a 3:1 ratio of yellow and green progeny.

(1) Mendel's Model of Particulate Genetics

* Mendel's first law predicts independent segregation of alleles at a single locus:
  * Two members of a gene pair (alleles) segregate separately into gametes so that half of the gametes carry one allele and the other half carry the other allele

(1) Mendel's Model of Particulate Genetics



Mendel's crosses to examine the segregation ratios of two phenotypes, seed coat (yellow or green) and seed coat surface (smooth or wrinkled) in pea plants. The hatched pattern indicates wrinkled seeds while white indicates smooth seeds. The F2 individuals exhibited a phenotypic ratio of 9 round/yellow : 3 round/green : 3 wrinkled/yellow : 1 wrinkled/green

(1) Mendel's Model of Particulate Genetics

* Mendel's second law predicts independent assortment of multiple loci:
    * during gamete formation, the segregation of alleles of one gene is independent of the segregation of alleles of another gene

(2) Hardy-Weinberg Expected Genotype Frequencies

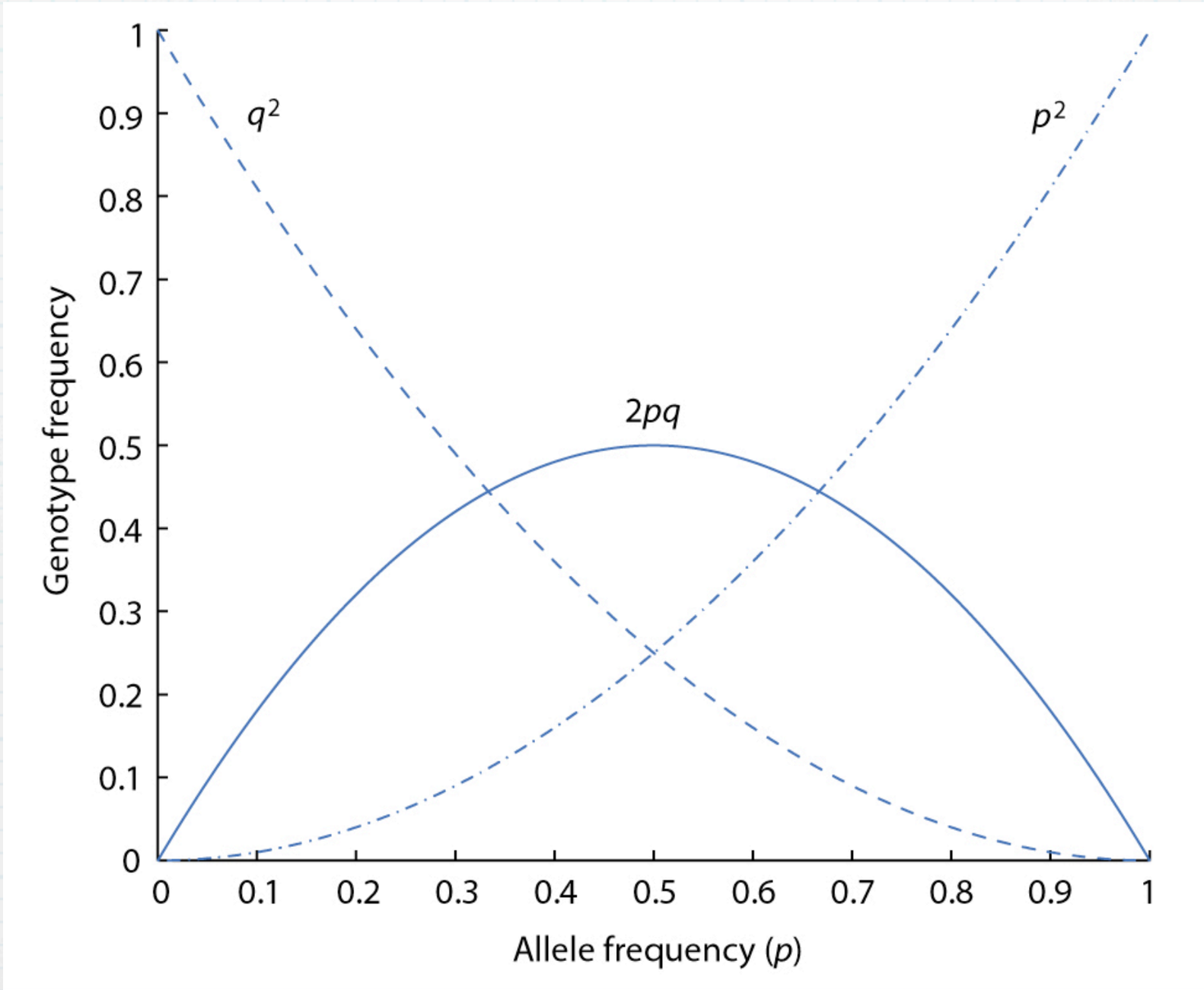
* In 1908, Hardy and Weinberg formulated the relationship that can be used to predict allele frequencies given genotype frequencies, or predict genotype frequencies given allele frequencies

* This relationship is the well-known Hardy-Weinberg equation

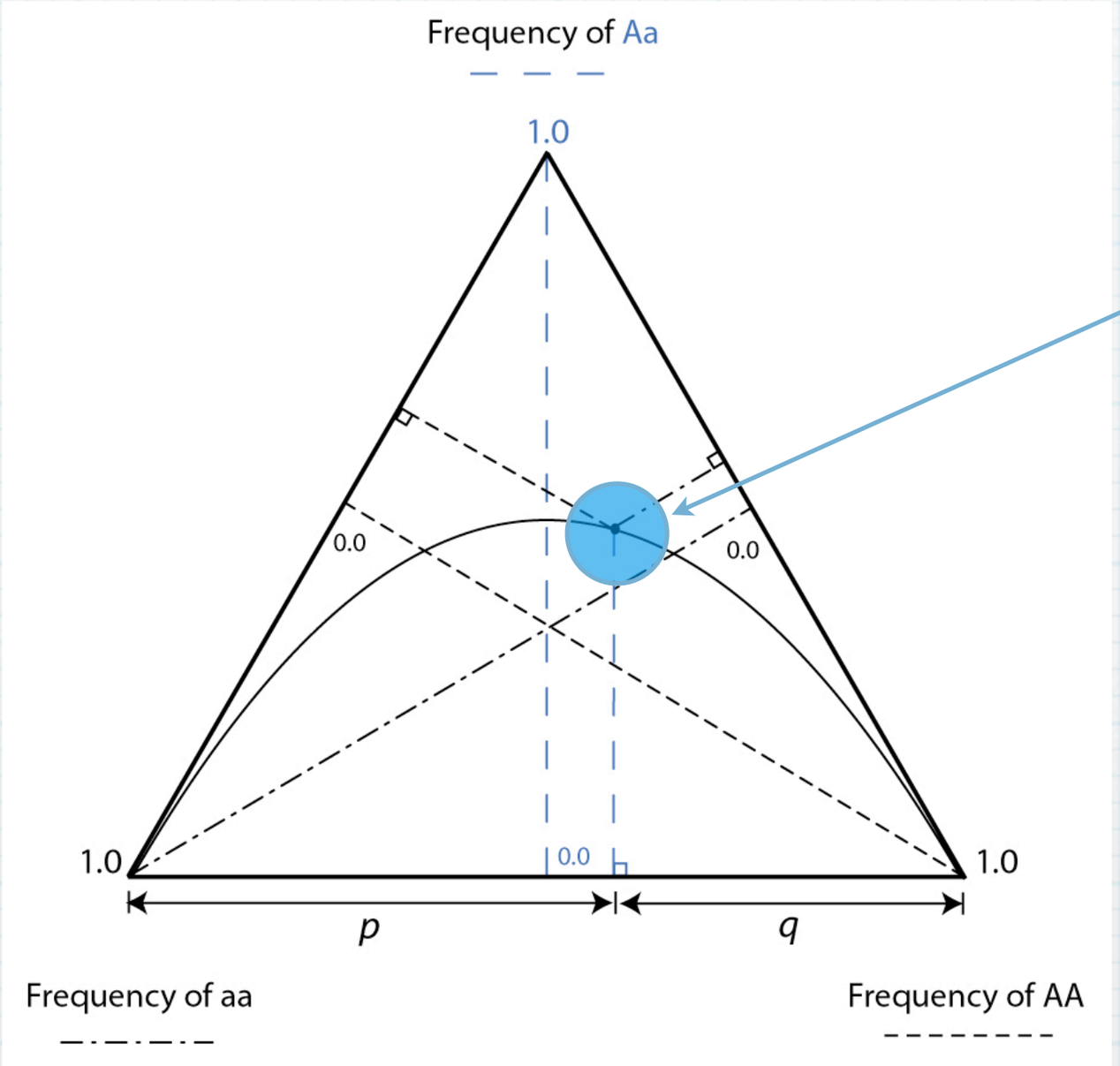
$$p^2+2pq+q^2=1$$

where p and q are allele frequencies for a locus with two alleles

(2) Hardy-Weinberg Expected Genotype Frequencies



(2) Hardy-Weinberg Expected Genotype Frequencies



A De Finetti diagram for one locus with two alleles.

(2) Hardy-Weinberg Expected Genotype Frequencies

* A single generation of reproduction where a set of conditions, or assumptions, are met will result in a population that meets Hardy-Weinberg expected genotype frequencies, often called Hardy-Weinberg equilibrium (HWE)
* The list includes:
  * the organism is diploid
  * reproduction is sexual (as opposed to clonal)

(2) Hardy-Weinberg Expected Genotype Frequencies

* discrete, non-overlapping generations
* the locus has two alleles
* allele frequencies are identical among mating types
* mating is random (as opposed to assortative)
* there is random union of gametes
* population size is effectively infinite

(2) Hardy-Weinberg Expected Genotype Frequencies

* migration is negligible (no population structure, no gene flow)
* mutation does not occur or its rate is very low
* natural selection does not act (all individuals and gametes have equal fitness)

(2) Hardy-Weinberg Expected Genotype Frequencies

* Question: So, what good is a model with so many restrictive assumptions that may not be met in actual populations?

* Answer:

  * HW provides a null model against which actual populations can be compared to test hypotheses about the evolutionary forces acting on allele and genotype frequencies.

  * Departure from HW expected genotype frequencies must be caused by processes that alter the outcome of basic inheritance

(2) Hardy-Weinberg Expected Genotype Frequencies

* One method for testing departure from HW is the χ^2 (chi-squared) test, which provides the probability of obtaining the difference between the observed and expected number of outcomes by chance alone if the null hypothesis is true

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

(2) Hardy-Weinberg Expected Genotype Frequencies

Example for testing departure from HW

* Observed genotypes for the MN blood group in a sample of 1066 individuals: 165 MM, 562 MN, and 339 NN

* First, we estimate the frequencies of the two alleles M and N

$$\hat{p} = \frac{2 \times \text{frequency(MM)} + \text{frequency(MN)}}{2N} = \frac{2 \times 165 + 562}{2 \times 1066} = 0.4184$$

$$\hat{q} = 1 - \hat{p} = 0.5816$$

(2) Hardy-Weinberg Expected Genotype Frequencies

Example for testing departure from HW

* Then, we compute the expected numbers of genotypes:
  * MM: $1066 \times (0.4184)^2 = 186.61$
  * MN: $1066 \times 2(0.4184)(0.5816) = 518.80$
  * NN: $1066 \times (0.5816)^2 = 360.58$
* Then, we compute the χ^2 statistic:

$$\chi^2 = \frac{(165 - 186.61)^2}{186.61} + \frac{(562 - 518.80)^2}{518.80} + \frac{(339 - 360.58)^2}{360.58} = 7.39$$

(2) Hardy-Weinberg Expected Genotype Frequencies

Example for testing departure from HW

* Finally, we need to compare the statistic to values from the chi-squared distribution for the right degrees of freedom (df)
  * df = (# classes compared) - (# parameters estimated) - 1
  * In our case, df = 3 - 1 - 1 = 1
  * A chi-squared value of 7.39 with 1 df has a probability between 0.01 and 0.001
  * Conclusion: the HW expected genotype frequencies are not present in the population

(2) Hardy-Weinberg Expected Genotype Frequencies

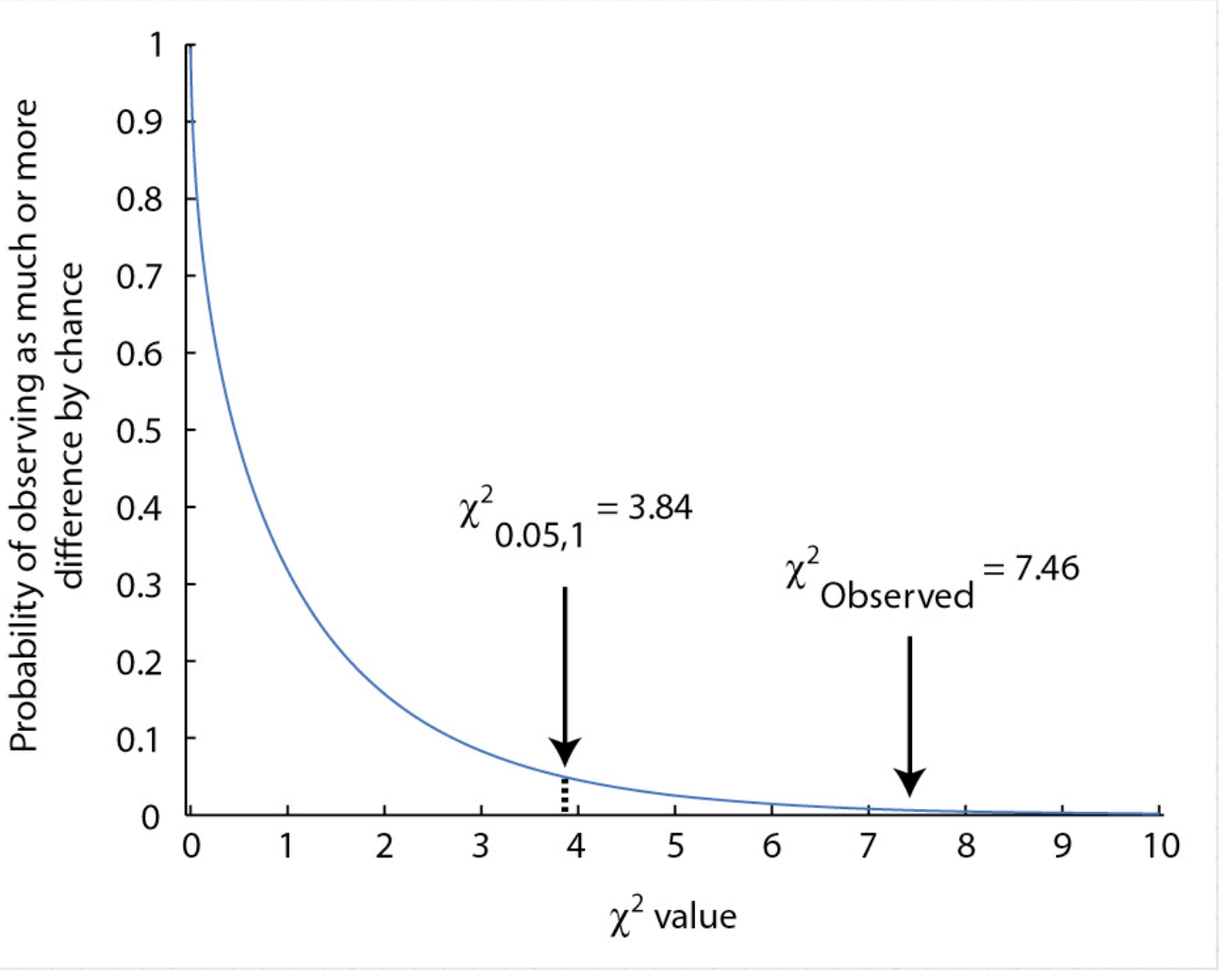
Example for testing departure from HW

Chi Square Distribution Table

d.f.	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	$\chi^2_{.001}$
1	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	22.7	27.2	30.1	32.9	36.2	38.6	32.8
20	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	28.2	33.2	36.4	39.4	32.0	45.6	51.2
25	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	77.6	85.5	90.5	95.0	100	104	112
80	88.1	96.6	102	107	112	116	125
90	98.6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

Table from Ronald J. Wonnacott and Thomas H. Wonnacott,
Statistics: Discovering its Power, New York: John Wiley and Sons, 1982, p.352.

By convention, we reject the null hypothesis if the chi-squared value has a probability of 0.05 or less



(2) Hardy-Weinberg Expected Genotype Frequencies

Example for testing departure from HW

* In a total of 3816 corn seeds, the following phenotypes were observed:

* Purple,smooth: 2058;
  Purple,wrinkled: 728;
  Yellow,smooth: 769;
  Yellow,wrinkled: 261

* Are these genotype frequencies consistent with inheritance due to **one locus with three alleles** or **two loci each with two alleles**?

(3) The Fixation Index and Heterozygosity

* Mating among related individuals, termed **consanguineous mating** or **biparental inbreeding**, increases the probability that the resulting progeny are homozygous compared to random mating

* The **fixation index** (F) is the proportion by which heterozygosity is reduced or increased relative to the heterozygosity in a randomly mating population with the same allele frequencies

(3) The Fixation Index and Heterozygosity

* If H_e is the HW expected frequency of heterozygotes based on population allele frequencies, and H_0 is the observed frequency of heterozygotes, then

$$F = \frac{H_e - H_o}{H_e}$$

(3) The Fixation Index and Heterozygosity

* Caution: Departures from the HW expected genotype frequencies estimated by F are potentially influenced by processes in addition to the mating system

* Genetic loci free of the influence of other processes, such as natural selection, are often sought to estimate F

* Further, F can be estimated using the average of multiple loci, which tends to reduce bias

(3) The Fixation Index and Heterozygosity

* For an arbitrary number of alleles at one locus, we compute H_e as

$$H_e = 1 - \sum_{i=1}^{k} p_i^2 = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 2p_i p_j$$

where k is the number of alleles at the locus, the p_i^2 and $2p_ip_j$ terms represent the expected genotype frequencies based on allele frequencies, and the summation is over the frequencies of the k homozygous genotypes

(3) The Fixation Index and Heterozygosity

* The observed heterozygosity Ho is estimated as

$$\hat{H}_o = \sum_{i=1}^{h} H_i$$

where the observed frequency of each heterozygous genotype H_i is summed over the h=k(k-1)/2 heterozygous genotypes possible with k alleles

* Finally, both H_e and H_o can be averaged over multiple loci to obtain mean heterozygosity estimates for two or more loci

(4) Mating Among Relatives

* **Positive assortative mating** describes the case when individuals with like genotypes or phenotypes tend to mate
* **Self-fertilization** is an extreme example of consanguineous mating where an individual can mate with itself by virtue of possessing reproductive organs of both sexes

(4) Mating Among Relatives

* The impact of complete positive genotypic assortative mating or self-fertilization on genotype frequencies

Generation	Genotype		
	AA	Aa	aa
0	D	H	R
	$H/4$	$\frac{1}{2}$	$H/4$
1	$D + H/4$	$H/2$	$R + H/4$
	$H/8$	$\frac{1}{2}$	$H/8$
2	$D + H/4 + H/8$	$H/4$	$R + H/4 + H/8$
t	$D + \frac{H}{2}[1 - (\frac{1}{2})^t]$	$H(\frac{1}{2})^t$	$R + \frac{H}{2}[1 - (\frac{1}{2})^t]$
infinity	$D + H/2$	0	$R + H/2$

(4) Mating Among Relatives

* In general, positive assortative mating or inbreeding changes the way in which alleles are "packaged" into genotypes, increasing the frequencies of all homozygous genotypes by the same total amount that heterozygosity is decreased, but allele frequencies in a population do not change

(4) Mating Among Relatives

Inbreeding coefficient and autozygosity

* The effects of consanguineous mating can also be thought of as increasing the probability that two alleles at one locus in an individual are inherited from the same ancestor
* Such a genotype would be homozygous and considered **autozygous**
* If the two alleles are not inherited from the same ancestor in the recent past, we would call the genotype **allozygous**

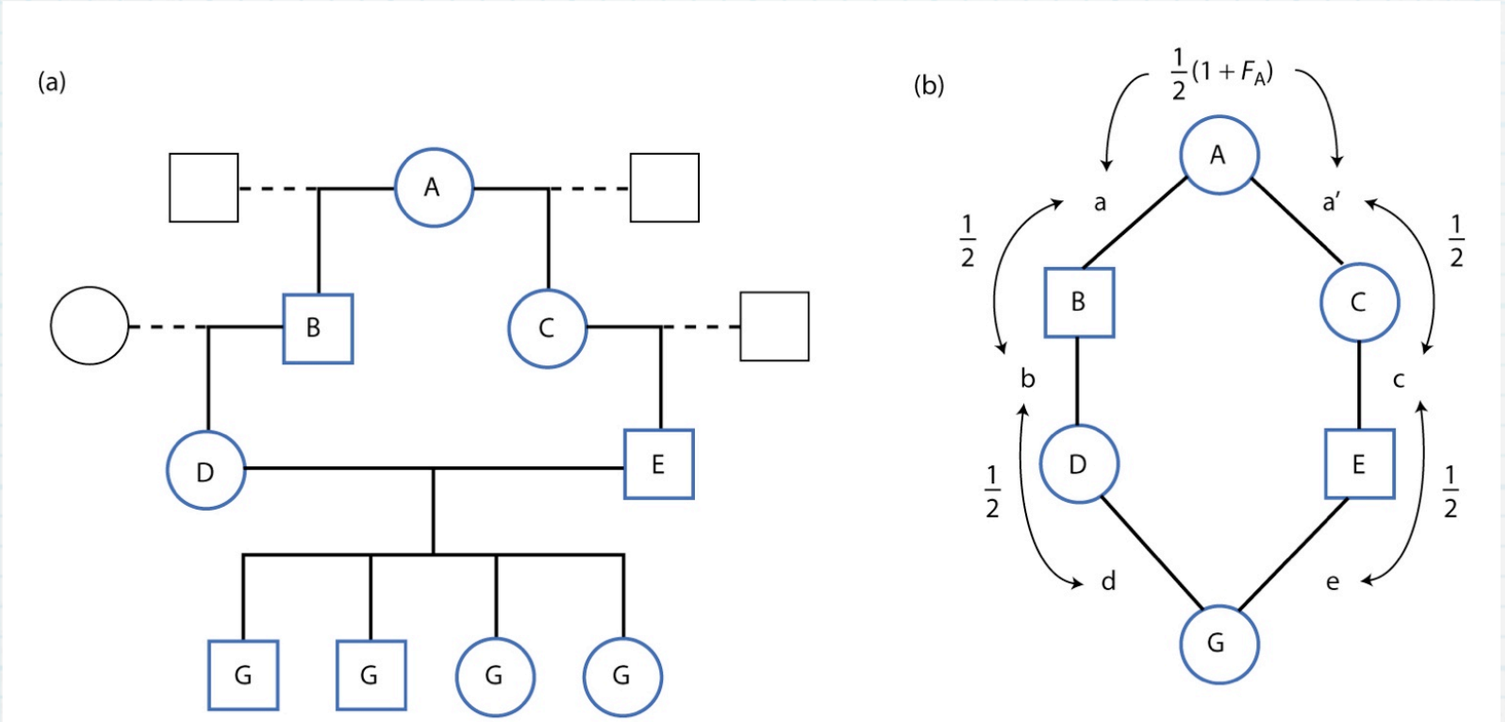
(4) Mating Among Relatives

Inbreeding coefficient and autozygosity

* Inbreeding in the autozygosity sense, often called **coefficient of inbreeding** (f), can be best seen in a **pedigree**

(4) Mating Among Relatives

Inbreeding coefficient and autozygosity



Average relatedness and autozygosity as the probability that two alleles at one locus are identical by descent

Path of inheritance	a	a′	a	a′	a	a′	a	a′
Allele	1	1	2	2	1	2	2	1
	Alleles identical by descent (IBD)				Alleles not identical by descent			

The possible patterns of transmission from one parent to two progeny for a locus with two alleles. Half of the outcome result in the two progeny inheriting an allele that is identical by descent

(4) Mating Among Relatives

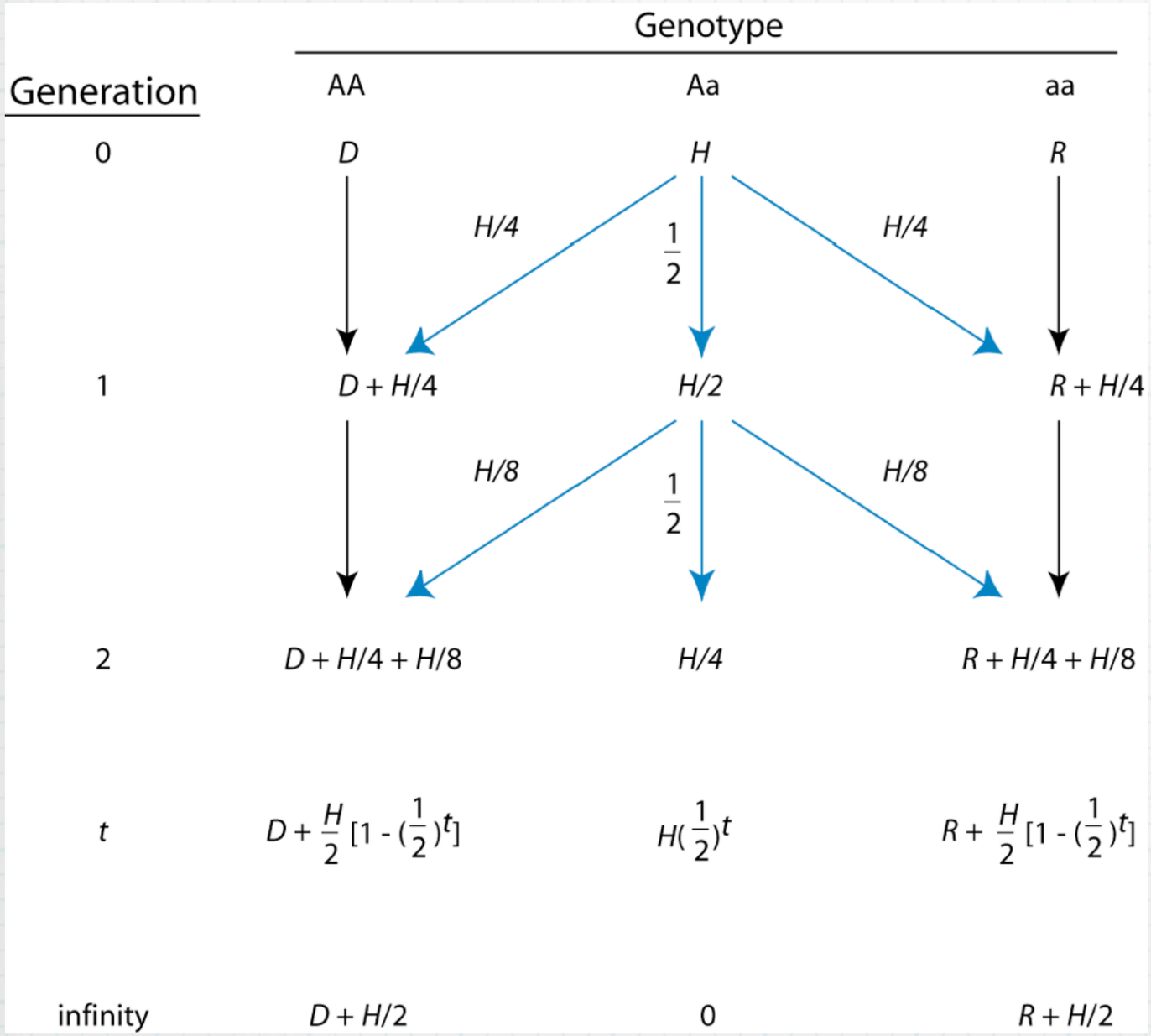
Inbreeding coefficient and autozygosity

* The departure from HW expected genotype frequencies, the autozygosity or inbreeding coefficient f, and the fixation index F, are all interrelated

* f measures the degree to which HW genotype proportions are not met, due to inbreeding

(4) Mating Among Relatives

Inbreeding coefficient and autozygosity



$$D=p^2+fpq$$
$$H=2pq-f2pq \quad (2.20)$$
$$R=q^2+fpq$$

Substituting He for 2pq

$$H=H_e(1-f)$$

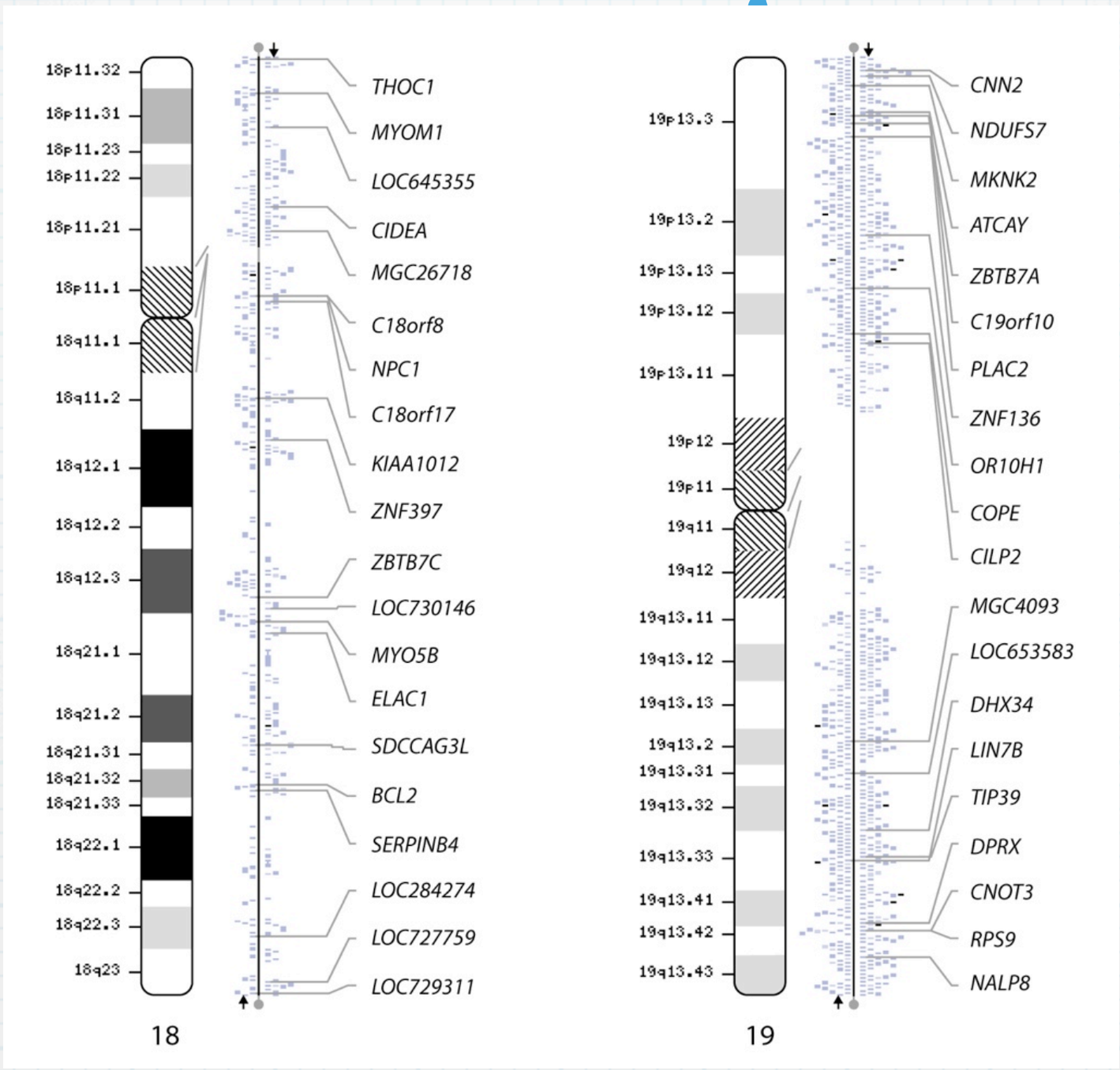
Rearranging

$$f=1-(H/H_e)=(H_e-H)/H_e$$

(5) Gametic Disequilibrium

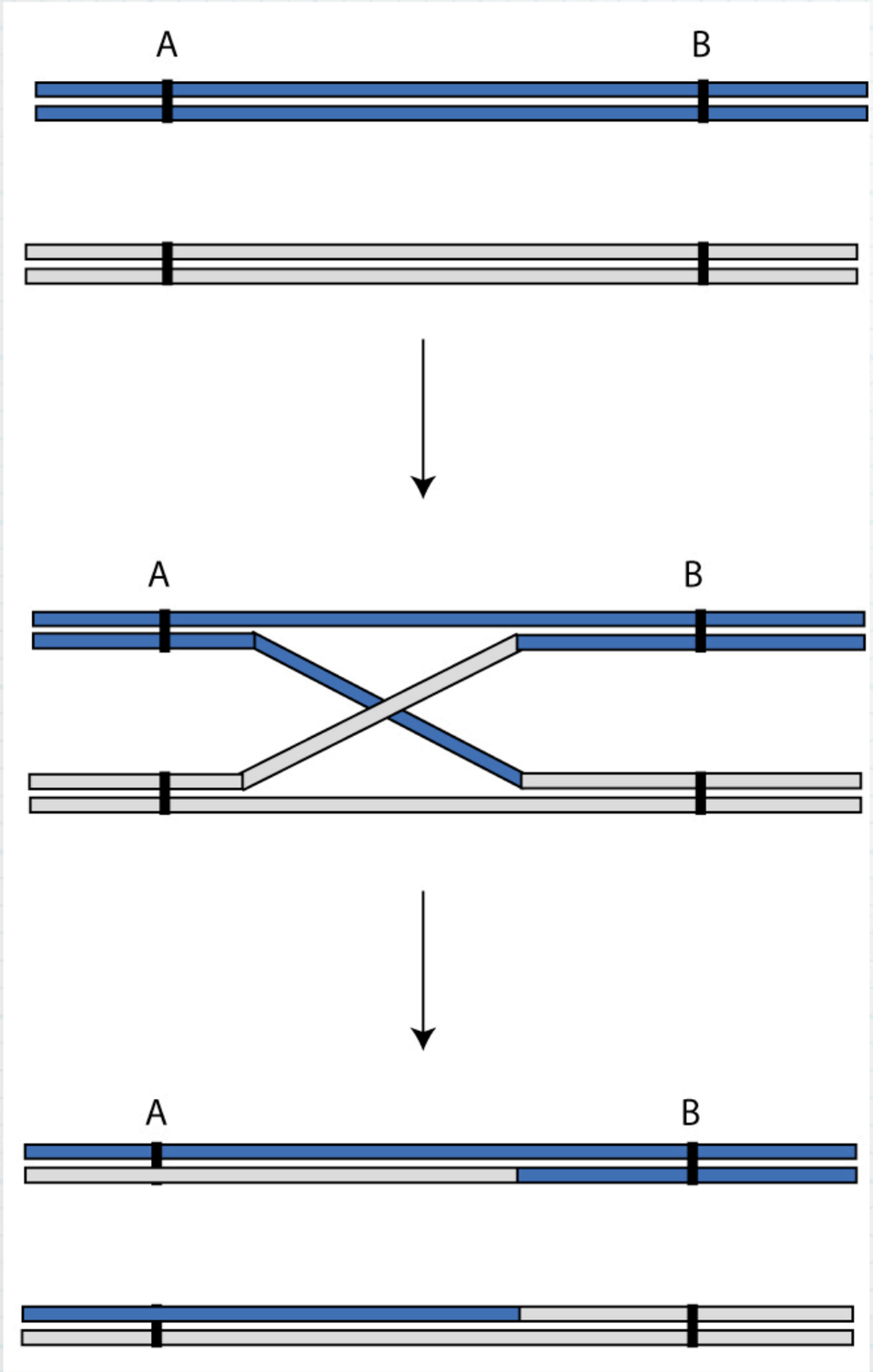
* Recall Mendel's second "law":
  * during gamete formation, the segregation of alleles of one gene is independent of the segregation of alleles of another gene
* This assumes the absence of linkage
* However, we know that some genes are physically **linked** by being on the same chromosome

(5) Gametic Disequilibrium



Maps for human chromosomes 18 and 19

(5) Gametic Disequilibrium



A schematic diagram of the process of recombination between two loci, A and B. Two double-stranded chromosomes exchange strands and form a Holliday structure. The **crossover** event can resolve into either of two recombinant chromosomes that generate new combinations of alleles at the two loci. **The chance of a crossover event occurring generally increases as the distance between loci increases.** Two loci are independent when the probability of recombination and non-recombination are both equal to 1/2.

(5) Gametic Disequilibrium

* To generalize expectations for genotype frequencies for two (or more) loci requires a model that accounts explicitly for linkage by including the rate of **recombination** between loci

* The frequency of a two-locus gamete haplotype depends on two factors: (i) allele frequencies, and (ii) the amount of recombination between the two loci

(5) Gametic Disequilibrium

* A model for two loci A (alleles A_1 and A_2, with frequencies p_1 and p_2) and B (alleles B_1 and B_2, with frequencies q_1 and q_2)

* The **recombination fraction**, symbolized as r, refers to the total frequency of gametes resulting from recombination events between two loci

* "Coupling" gametes: A_1B_1 and A_2B_2

* "Repulsion" gametes: A_1B_2 and A_2B_1

(5) Gametic Disequilibrium

$$\frac{A_1 \quad B_1}{A_2 \quad B_2}$$

Gamete	Frequency		
	Expected	Observed	
A_1B_1	p_1q_1	$g_{11}=(1-r)/2$	1-r is the frequency of all coupling gametes
A_2B_2	p_2q_2	$g_{22}=(1-r)/2$	
A_1B_2	p_1q_2	$g_{12}=r/2$	r is the frequency of all recombinant gametes
A_2B_1	p_2q_1	$g_{21}=r/2$	

(5) Gametic Disequilibrium

Expected frequencies of gametes for two diallelic loci in a randomly mating population with a recombination rate between the loci of r

		Frequency of gametes in next generation			
Genotype	Expected frequency	A_1B_1	A_2B_2	A_1B_2	A_2B_1
A_1B_1/A_1B_1	$(p_1q_1)^2$	$(p_1q_1)^2$			
A_2B_2/A_2B_2	$(p_2q_2)^2$		$(p_2q_2)^2$		
A_1B_1/A_1B_2	$2(p_1q_1)\ (p_1q_2)$	$(p_1q_1)\ (p_1q_2)$		$(p_1q_1)\ (p_1q_2)$	
A_1B_1/A_2B_1	$2(p_1q_1)\ (p_2q_1)$	$(p_1q_1)\ (p_2q_1)$			$(p_1q_1)\ (p_2q_1)$
A_2B_2/A_1B_2	$2(p_2q_2)\ (p_1q_2)$		$(p_2q_2)\ (p_1q_2)$	$(p_2q_2)\ (p_1q_2)$	
A_2B_2/A_2B_1	$2(p_2q_2)\ (p_2q_1)$		$(p_2q_2)\ (p_2q_1)$		$(p_2q_2)\ (p_2q_1)$
A_1B_2/A_1B_2	$(p_1q_2)^2$			$(p_1q_2)^2$	
A_2B_1/A_2B_1	$(p_2q_1)^2$				$(p_2q_1)^2$
A_2B_2/A_1B_1	$2(p_2q_2)\ (p_1q_1)$	$(1-r)(p_2q_2)\ (p_1q_1)$	$(1-r)(p_2q_2)\ (p_1q_1)$	$r(p_2q_2)\ (p_1q_1)$	$r(p_2q_2)\ (p_1q_1)$
A_1B_2/A_2B_1	$2(p_1q_2)\ (p_2q_1)$	$r(p_1q_2)\ (p_2q_1)$	$r(p_1q_2)\ (p_2q_1)$	$(1-r)(p_1q_2)\ (p_2q_1)$	$(1-r)(p_1q_2)\ (p_2q_1)$

(5) Gametic Disequilibrium

* We can utilize observed gamete frequencies to develop a measure of the degree to which alleles are associated within gamete haplotypes

* This quantity is called the **gametic disequilibrium**, or **linkage disequilibrium**, parameter, and can be expressed by:

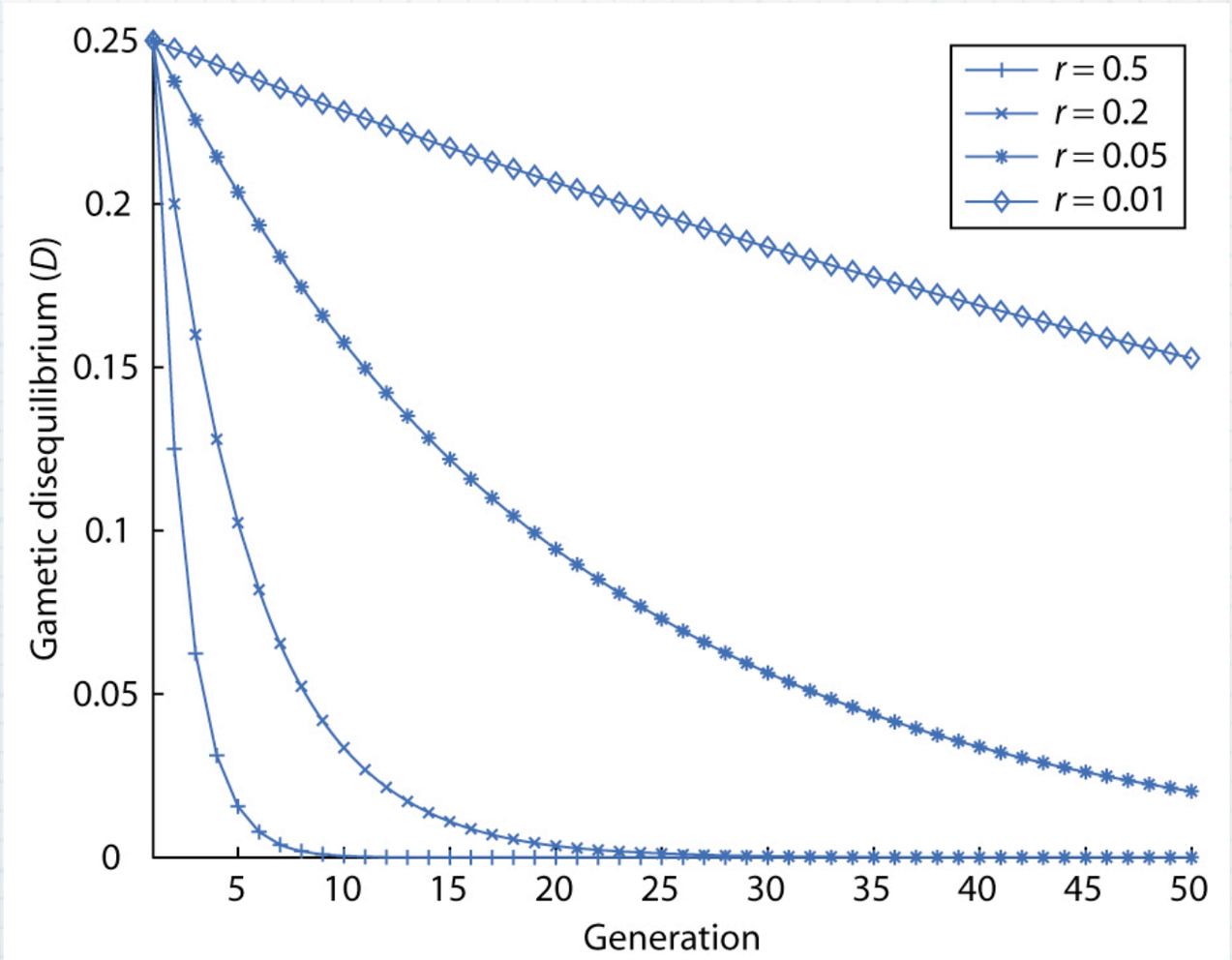
$$D = \underbrace{g_{11}g_{22}}_{\text{(coupling term)}} - \underbrace{g_{12}g_{21}}_{\text{(repulsion term)}}$$

* One estimator of gametic disequilibrium is the squared correlation coefficient

$$\rho^2 = D^2/(p_1 p_2 q_1 q_2)$$

(5) Gametic Disequilibrium

* How does D change over time?
* Recombination produces r recombinant gametes each generation, so that $D_t=(1-r)D_{t-1}$
* Since gametic disequilibrium decays by a factor of 1-r each generation, we have $D_t=D_0(1-r)^t$



Initially, there are only coupling ($P_{11}=P_{22}=1/2$) and no repulsion gametes ($P_{12}=P_{21}=0$)

(5) Gametic Disequilibrium

* D depends on the allelic frequencies in the population

* This can make interpreting or comparing D estimates problematic

* A way to avoid these problems is to express D as the percentage of its largest value $D' = D/D_{max}$, where D_{max} is the larger of $-p_1q_1$ and $-p_2q_2$ when $D<0$, and the smaller of p_1q_2 and p_2q_1 when $D>0$

(5) Gametic Disequilibrium

Physical Linkage

* Linkage is the physical association of loci on a chromosome that causes alleles at the loci to be inherited in their original combinations

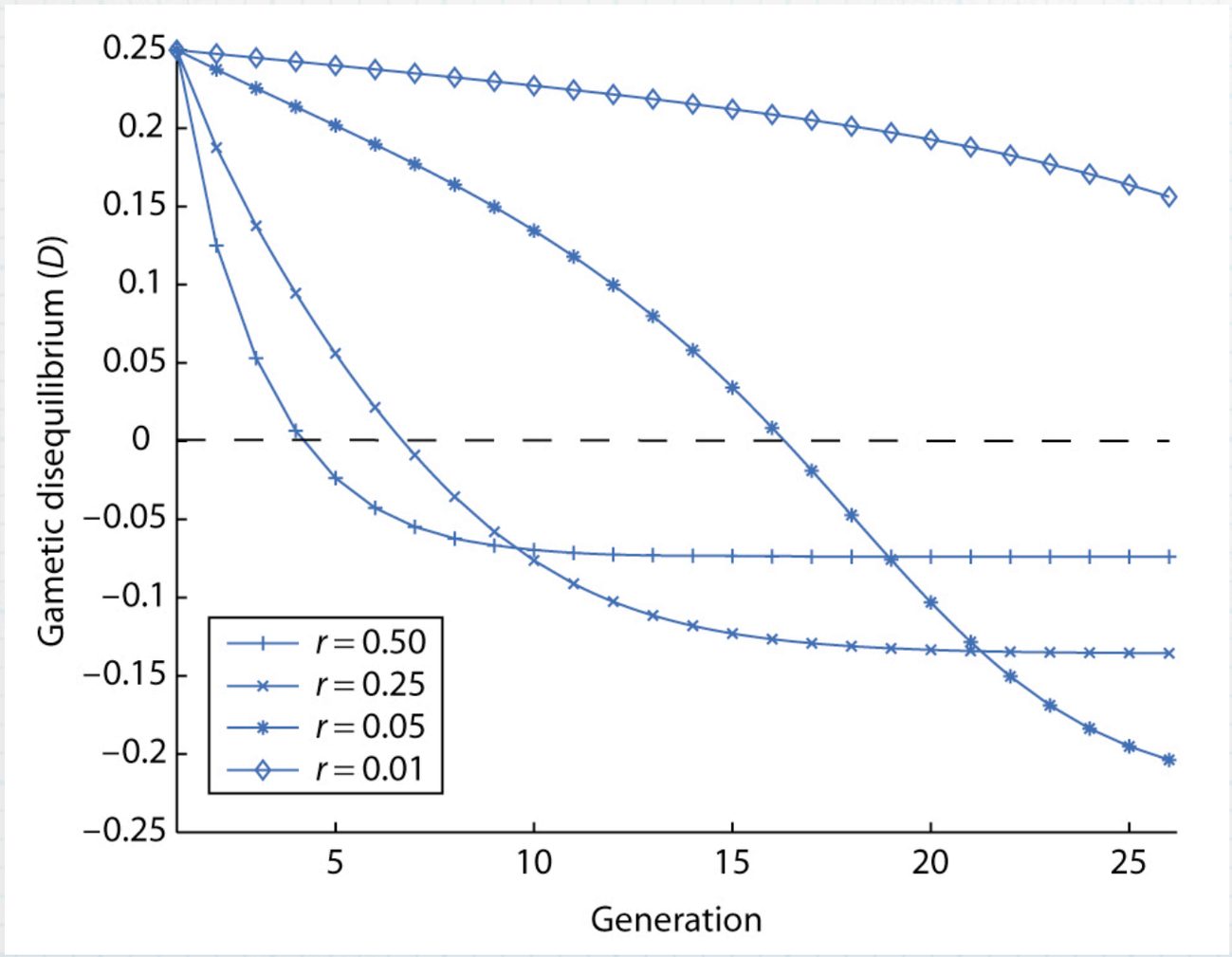
* The degree of physical linkage of loci dictates the recombination rate and thereby the decay of gametic disequilibrium

* Linkage-like effects can be seen in some chromosomes and genomes where gametic disequilibrium is expected to persist over longer time scales due to exceptional inheritance or recombination patterns (e.g., organelle genomes)

(5) Gametic Disequilibrium

Natural Selection

* Natural selection can continuously counteract the randomizing effects of recombination, potentially producing steady states other than D=0



- Initially, there are only coupling ($P_{11}=P_{22}=1/2$) and no repulsion gametes ($P_{12}=P_{21}=0$)
- The relative fitness values of the AAbb and aaBB genotypes are 1, while all other genotypes have a fitness of 1/2
- Gametic disequilibrium does not decay to 0 over time due to the action of natural selection

(5) Gametic Disequilibrium

Mutation

* Alleles change from one form to another by the random process of mutation, which can either increase or decrease gametic disequilibrium

* A novel allele produced by mutation would initially increase gametic disequilibrium, and recombination then woks to randomize the other alleles found with the novel allele, and eventually dissipate the gametic disequilibrium

* Mutation can also produce alleles identical to those currently present in the population. In this case, mutation can contribute to randomizing the combinations of alleles at different loci and thereby decrease levels of gametic disequilibrium

(5) Gametic Disequilibrium

Mutation

* It is important to recognize that mutation rates are often very low and the gamete frequency changes caused by mutation are inversely proportional to population size, so that mutation usually makes a modest contribution to overall levels of gametic disequilibrium

(5) Gametic Disequilibrium

Mixing of Diverged Populations

* The mixing of two diverged populations, often termed admixture, can produce substantial levels of gametic disequilibrium

* This is caused by different allele frequencies in the two source populations that result in different gamete frequencies at gametic disequilibrium

* In general, gametic disequilibrium due to admixture increases as allele frequencies become more diverged between the source populations, and the initial composition of the mixture population approaches equal proportions of the source populations

(5) Gametic Disequilibrium

Mating System

* Self-fertilization and mating between relatives increases homozygosity (at the expense of heterozygosity)

* An increase in homozygosity causes a reduction in the effective rate of recombination because crossing over between two homozygous loci does not alter the gamete haplotypes produced by that genotype

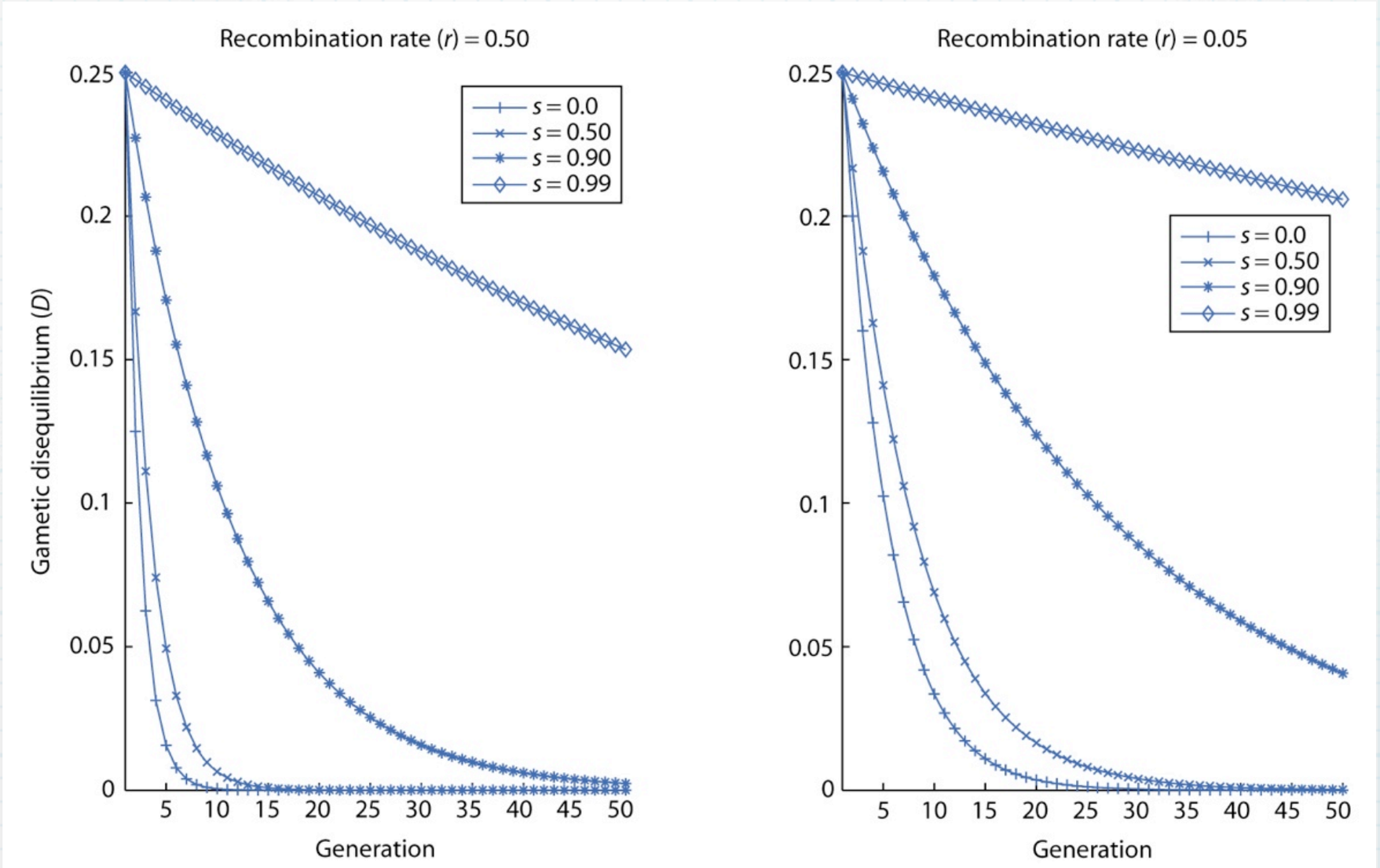
* The effective recombination fraction under self-fertilization is:

$$r_{effective} = r\left(1 - \frac{s}{2 - s}\right)$$

where s is the proportion of progeny produced by self-fertilization in each generation

(5) Gametic Disequilibrium

Mating System



Initially, there are only coupling (P_{11}=P_{22}=1/2) and no repulsion gametes (P_{12}=P_{21}=0)

(5) Gametic Disequilibrium

Chance

* It is possible to observe gametic disequilibrium just by chance in small populations or small samples of gametes

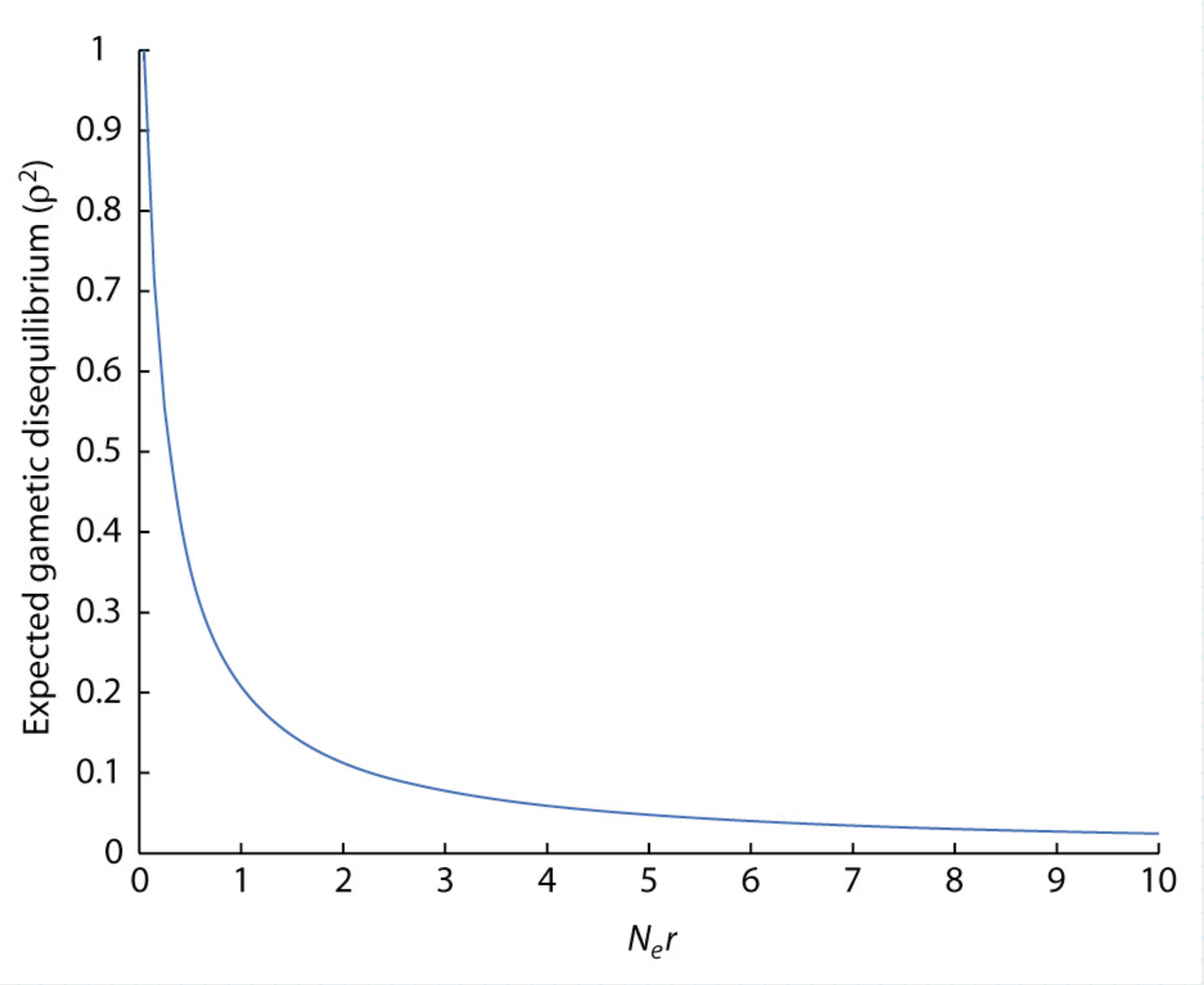
* When the chance effects due to population size and recombination are in equilibrium, the effects of population size can be summarized approximately by

$$\rho^2 = \frac{1}{1 + 4N_e r}$$

where N_e is the genetic effective population size, and r is the recombination fraction per generation

(5) Gametic Disequilibrium

Chance



Summary

* Hardy-Weinberg expected genotype frequencies serve as a null model used as a standard reference
* The fixation index (F) measures departure from HW caused by patterns of mating
* Consanguineous mating changes genotype, but not allele, frequencies
* Consanguineous mating can be viewed as a process that increases that chances that alleles descended from a common ancestor are found together in a diploid genotype (autozygosity)
* The fixation index, autozygosity, and coefficient of inbreeding are all interrelated measures of changes in genotype frequencies with consanguineous mating

Summary

* Consanguineous mating may result in inbreeding depression, which ultimately is caused by over-dominance (heterozygote advantage) or dominance (deleterious recessive alleles)
* The gametic disequilibrium parameter (D) measures the degree of non-random association of alleles at two loci
* Gametic disequilibrium is broken down by recombination
* A wide variety of population genetic processes (natural selection, chance, admixture, mating system, and mutation) can maintain and increase gametic disequilibrium even between loci without physical linkage to reduce recombination