

# Basic Principles of Forensic Molecular Biology and Genetics



## Population Genetics

# Significance of a Match



What is the significance of:

- a fiber match?
- a hair match?
- a glass match?
- a DNA match?

# Meaning of a DNA Match

- Approximately 1 person in every 5 trillion chosen at random from the population would be expected to possess the same DNA genotype as that found in the semen
- or
- The DNA results are 5 trillion times more likely if the semen originated from the suspect than if it had originated from a randomly chosen unrelated individual from the population

# Meaning of a DNA Match (cont'd)

- Must be able to estimate the frequency of occurrence of the DNA genotype in the relevant population
- Need to study population genetics
  - population genetics is concerned with how much genetic variation exists in natural populations and explains its origin, maintenance and evolutionary importance
    - genetic variation: no two individuals would be expected to possess the same genotype for all genes

# **Population Genetics Lectures**



- Organization of Genetic Variation
- Population Substructure
- Forensic Genetic Data Analysis

# Some Definitions

- genotype: genetic constitution of an individual
- phenotype: physical or biochemical expression of a genotype ('traits')
- polymorphism: a gene (or DNA locus) for which the most common allele has a frequency of  $< 0.95$  (or  $0.99$ )
  - | c.f. rare allele (frequency  $< 0.005$ )
  - | With 2 allele polymorphism, at least 9.5% of population would be heterozygous for the common allele

# Allele Frequency Determination

- Allele frequency: proportion of all alleles at a locus of a particular type
  - frequency of an allele =  $2 \times \text{No. of homozygotes (because 2 copies of the allele)} + \text{No. of heterozygotes possessing the allele (because heterozygotes contain 1 copy of the allele)}$  divided by  $2 \times \text{No. of individuals (because each individual carries 2 alleles at each locus)}$

# Allele Frequency Determination

- Three genotypes: AA, Aa, aa  
 $p_A$  = frequency of allele A  
 $p_a$  = frequency of allele a
- P = proportion of AA individuals, Q = proportion of Aa, R = proportion of aa
- $P + Q + R = 1$
- $p_A = (2n_{AA} + n_{Aa})/2n$ , or
- $p_A = P + Q/2$



# Random Mating

Genotypes are not transmitted from one generation to the next

- broken up in gamete formation by segregation and recombination and assembled anew in each generation by fertilization
  - genotypes  $\rightarrow$  gametes  $\rightarrow$  genotypes
- Model: mating takes place at random with respect to the genotypes under consideration
  - If AA 0.16, Aa 0.48, aa 0.36 then AA males mate with AA, Aa and aa females in these proportions

# Populations

- Group of individuals of the same species living within a restricted geographical location such that any member can mate with any other member
  - local interbreeding units within large geographical structures
    - local populations / demes or sub-populations
- Non overlapping generation model
  - birth/maturation/reproduction/death in generation  $t-1$  followed by  $t$  followed by  $t+1$

# Hardy – Weinberg Model

- Genotype frequencies expected due to random mating (Hardy and Weinberg independently described in 1908)
- Assumptions
  - diploid organism
  - sexual reproduction
  - non overlapping generations
  - mating is random
  - population size is large
  - migration/mutation/selection insignificant

# Hardy – Weinberg Principle

- With 2 alleles of a gene (A and a with frequencies  $p$  and  $q$ ) there are 6 possible matings AA/AA, AA/Aa, AA/aa, Aa/Aa, Aa/aa, aa/aa
  - genotype frequencies are AA: $p^2$ , Aa:  $2pq$ , aa: $q^2$
- random mating of individuals is usually equivalent to the random union of gametes

Table 3. Demonstration of the Hardy–Weinberg principle.

Mating	Frequency of mating	OFFSPRING GENOTYPE FREQUENCIES		
		AA	Aa	aa
AA × AA	$P^2$	1	0	0
AA × Aa	$2PQ$	$\frac{1}{2}$	$\frac{1}{2}$	0
AA × aa	$2PR$	0	1	0
Aa × Aa	$Q^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Aa × aa	$2QR$	0	$\frac{1}{2}$	$\frac{1}{2}$
aa × aa	$R^2$	0	0	1
Totals (next generation)		$\frac{P'}{P'}$	$\frac{Q'}{Q'}$	$\frac{R'}{R'}$

where  $P' = P^2 + 2PQ/2 + Q^2/4 = (P + Q/2)^2 = p^2$

$Q' = 2PQ/2 + 2PR + Q^2/2 + 2QR/2 = 2(P + Q/2)(R + Q/2) = 2pq$

$R' = Q^2/4 + 2QR/2 + R^2 = (R + Q/2)^2 = q^2$

# Hardy – Weinberg Principle

		Sperm	
		$A(p)$	$a(q)$
Eggs	$A(p)$	$AA(p^2)$	$Aa(pq)$
	$a(q)$	$Aa(pq)$	$aa(q^2)$

Frequencies in offspring:

$$AA:P' = p^2$$
$$Aa:Q' = pq + pq = 2pq$$
$$aa:R' = q^2$$

Figure 11. Cross-multiplication square showing Hardy–Weinberg frequencies resulting from random mating with two alleles.

# Hardy – Weinberg Principle

	$A_1(p_1)$	$A_2(p_2)$	$A_3(p_3)$
Eggs $A_1(p_1)$	$A_1A_1(p_1^2)$	$A_1A_2(p_1p_2)$	$A_1A_3(p_1p_3)$
$A_2(p_2)$	$A_1A_2(p_1p_2)$	$A_2A_2(p_2^2)$	$A_2A_3(p_2p_3)$
$A_3(p_3)$	$A_1A_3(p_1p_3)$	$A_2A_3(p_2p_3)$	$A_3A_3(p_3^2)$

Frequencies in offspring:

- $A_1A_1: p_1^2$
- $A_1A_2: p_1p_2 + p_1p_2 = 2p_1p_2$
- $A_1A_3: p_1p_3 + p_1p_3 = 2p_1p_3$
- $A_2A_2: p_2^2$
- $A_2A_3: p_2p_3 + p_2p_3 = 2p_2p_3$
- $A_3A_3: p_3^2$

Figure 14. Cross-multiplication square showing Hardy–Weinberg frequencies for three autosomal alleles.

# Hardy – Weinberg Principle



- Constancy of allele frequencies from generation to generation
  - therefore, in any generation genotype frequencies are  $p^2$ ,  $2pq$  and  $q^2$
  - Hardy Weinberg equilibrium



# Mating Systems

Table 4. Characteristics of several mating systems.

Mating system	Defining feature
Random mating	Choice of mates independent of genotype and phenotype
Positive assortative mating	Mates phenotypically more similar than would be expected by chance
Negative assortative mating (disassortative)	Mates phenotypically more dissimilar than would be expected by chance
Inbreeding	Mating between relatives

# Variance in Allele Frequency Estimates

3 alleles  $A_1, A_2, A_3$

$P_{ii}$  = proportion of  $A_iA_i$  homozygotes

$P_{ij}$  = proportion of  $A_iA_j$  heterozygotes

[Allele proportions  $p_i = P_{11} + 1/2P_{12} + 1/2P_{13}$ ]

$p_i = P_{ii} + 1/2 \sum P_{ij}$        $\text{Var}(p) = p(1-p)/2n$

Standard (std) error =  $[\text{var}(p)]^{1/2}$

68% +/- 1 std error, 95% +/- 2 std errors, 99.7% +/- 3 std errors

# Product Rule



$$\text{Product rule} = \prod p_i^2 \times \prod 2p_i p_j$$

# Inbreeding

Mating between relatives

assume  $\frac{1}{4}$  AA,  $\frac{1}{2}$  Aa,  $\frac{1}{4}$  aa

selfing (1st generation)  $\frac{3}{8}$  AA,  $\frac{2}{8}$  Aa,  $\frac{3}{8}$  aa

selfing (2nd generation)  $\frac{7}{16}$  AA,  $\frac{2}{16}$  Aa,  $\frac{7}{16}$  aa

REDUCTION IN HETEROZYGOSITY

( $\frac{4}{8} \rightarrow \frac{3}{8} \rightarrow \frac{1}{8}$ )

# **F = inbreeding coefficient**

F = fractional reduction in heterozygosity relative to random mating population

$$F = (H_0 - H) / H_0 \quad (\text{or } H = H_0 - H_0 F)$$

where  $H_0$  = random mating heterozygosity

and  $H$  = actual frequency of heterozygotes in inbreeding population

but since  $H_0 = 2pq$ , then  $H = H_0 - H_0 F = H_0(1-F)$

$$\text{or } H = 2pq(1-F)$$

# Genotype Proportions with Inbreeding

$$\text{Heterozygotes} = 2pq(1-F)$$

Homozygotes

Suppose proportion of AA is P and p = allele frequency of A  
 then  $P + H/2 = p$  or  $P = p - H/2$  but  $H = 2pq(1-F)$

$$\begin{aligned} \text{Therefore, } P &= p - 2pq(1-F)/2 = p - pq(1-F) = p - p(1-p)(1-F) \\ &= p - (p - p^2)(1-F) = p - [p - pF - p^2 + p^2F] = p - p + pF + p^2 - p^2F = pF + \\ p^2 - p^2F &= \boxed{pF + p^2(1-F)} = p^2 - p^2F + pF = p^2 + pF - p^2F = p^2 \\ &+ pF(1-p) \end{aligned}$$

$$AA = pF + p^2(1-F) = p^2 + pF(1-p)$$

$$Aa = 2pq(1-F)$$

$$aa = qF + q^2(1-F) = q^2 + qF(1-q)$$

If  $F = 0$  then Hardy Weiberg proportions pertain

If  $F = 1$  then complete inbreeding (only AA and aa types)

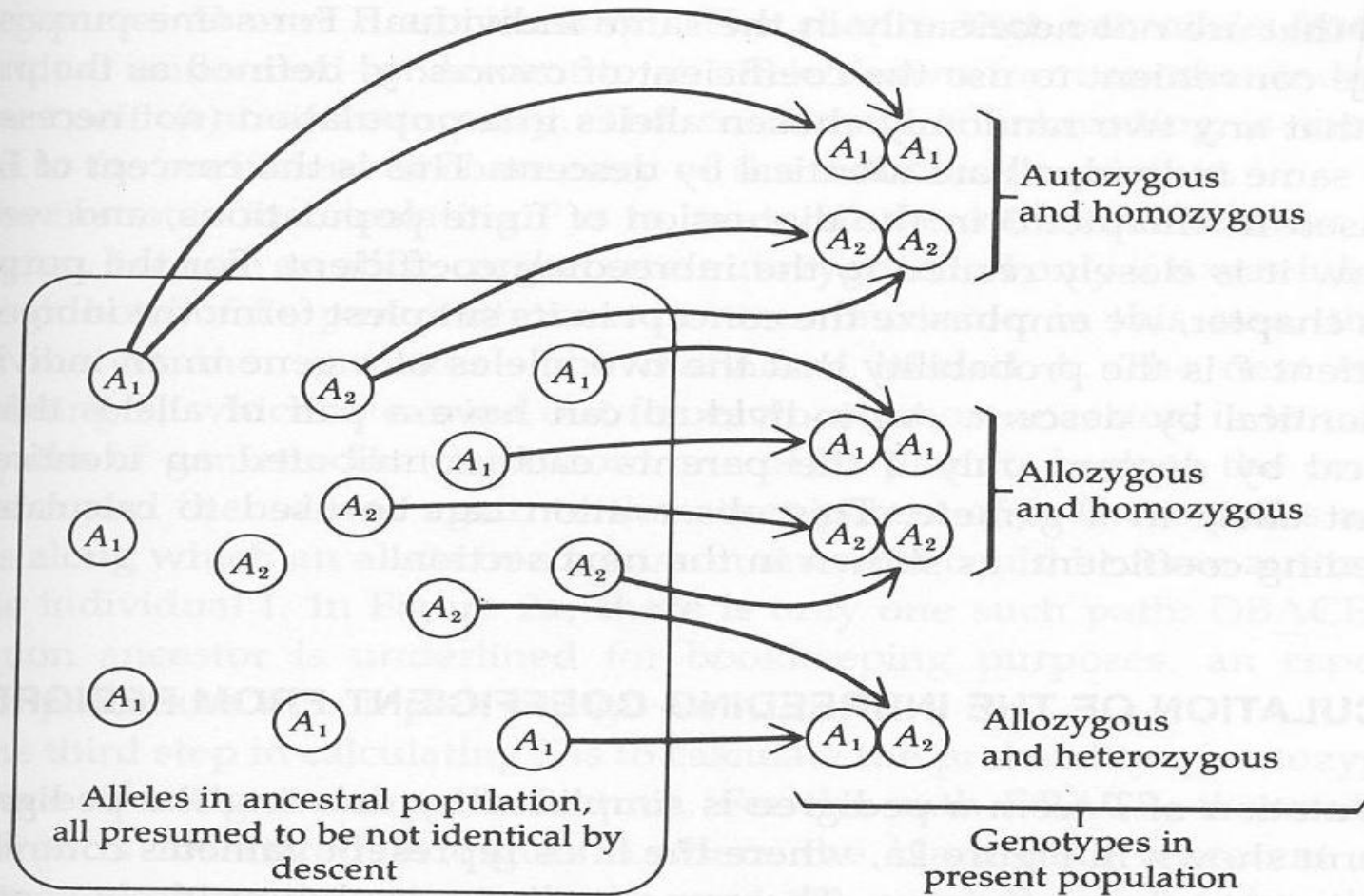
# Genotype Proportions with Inbreeding (cont'd)

If population contains multiple alleles  $A_1, A_2, A_3 \dots A_n$   
at frequencies  $p_1, p_2, p_3 \dots p_n$  ( with  $p_1 + p_2 + p_3 \dots + p_n = 1$  )

then in a population with inbreeding coefficient  $F$   
the frequencies of  $A_i A_i$  homozygotes and  $A_i A_j$  heterozygotes  
are:

$$A_i A_i = p_i F + p_i^2 (1-F)$$

$$A_i A_j = 2p_i q_j (1-F)$$



**Figure 1.** In an autozygous individual, homologous alleles are derived from replication of a single DNA sequence in an ancestor, and they are therefore identical by descent. In an allozygous individual, homologous alleles are not identical by descent. As shown here, allozygous individuals may be heterozygous or homozygous, but autozygous individuals must be homozygous (except in the unlikely event that one allele undergoes a mutation).



# Alternative definition of F

- F is the probability that the two alleles of a gene in a single individual are identical by descent (ibd) (autozygous)
  - ibd = autozygous
  - non-ibd = allozygous
- ibd refers to ancestral origin and not chemical makeup
  - autozygous individuals must be homozygous
  - allozygous individuals may be homozygous or heterozygous

# Equivalence of the Autozygosity and Heterozygosity Definitions of F

consider a population with an average inbreeding coefficient F

consider the alleles of one individual, either

1. Alleles allozygous ( $Pr = 1-F$ ) or
2. Alleles autozygous ( $Pr = F$ )

If allozygous then  $Pr(\text{genotype}) = \text{probability in random mating population (because no inbreeding)}$

If autozygous then  $Pr(\text{genotype}) = Pr(\text{allele frequency})$  since knowing what allele is present determines that the same one is present in the other copy

# Equivalence of the Autozygosity and Heterozygosity Definitions of F (cont'd)



Consider 2 alleles A and a, with frequencies p and q and  $p + q = 1$

$$\Pr(AA) = p^2(1-F) \text{ [allozygous]} + pF \text{ [autozygous]}$$

$$\Pr(Aa) = 2pq(1-F) \text{ [allozygous]}$$

# Genotype Frequencies with Inbreeding

Table 5. Genotype frequencies with inbreeding.

Genotype	FREQUENCY IN POPULATION		
	With inbreeding coefficient $F$	With $F = 0$ (random mating)	With $F = 1$ (complete inbreeding)
$AA$	$p^2(1 - F) + pF$	$p^2$	$p$
$Aa$	$2pq(1 - F)$	$2pq$	$0$
$aa$	$q^2(1 - F) + qF$	$q^2$	$q$
	$\underbrace{\hspace{10em}}$ Allozygous genes	$\underbrace{\hspace{10em}}$ Autozygous genes	

# Genetic Drift

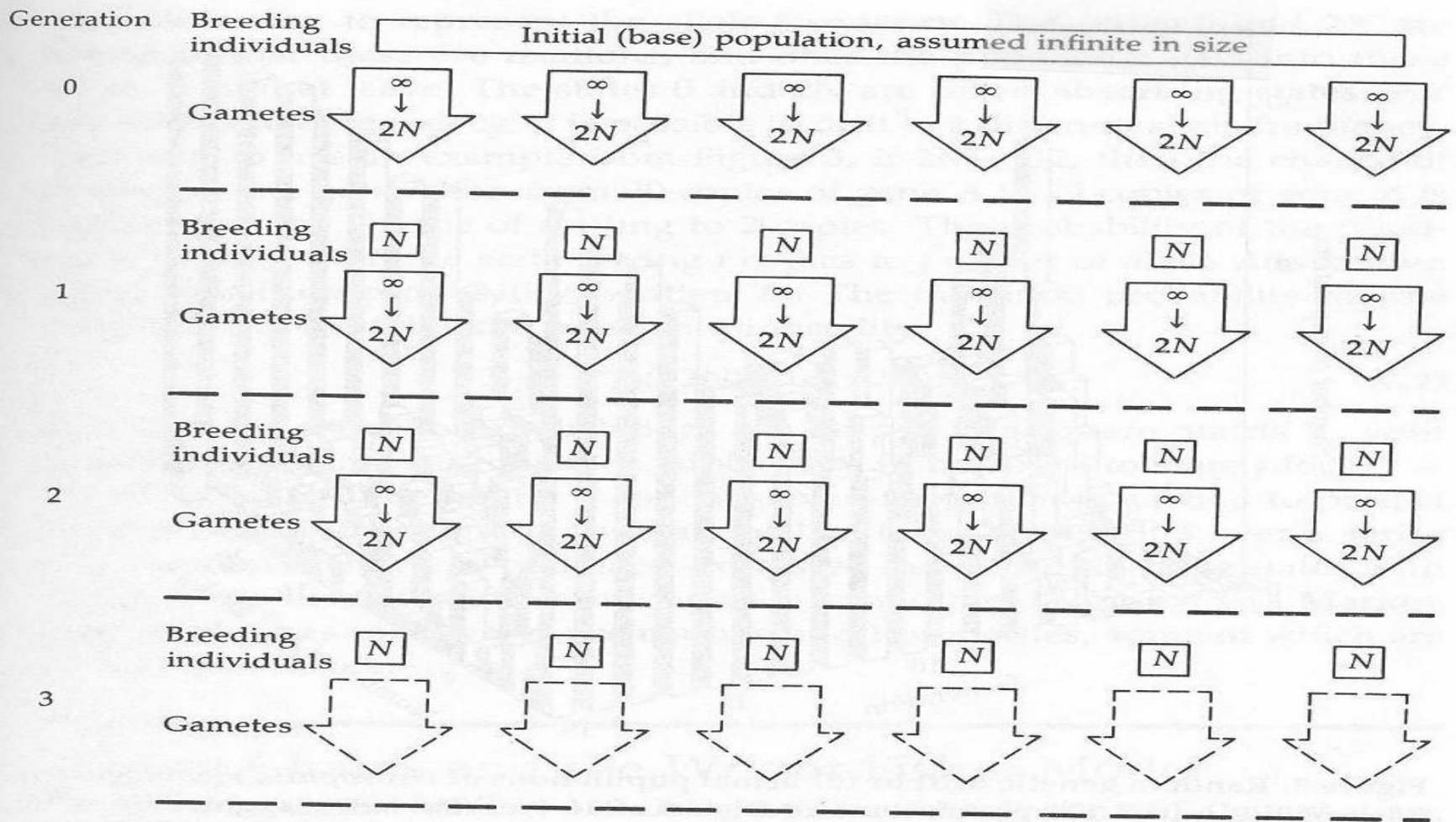


- Chance fluctuations in allele frequencies which occur, particularly in small populations, as a result of random sampling of gametes
  - Because small samples are frequently not representative, an allelic frequency in the sample may differ from that in the entire pool of gametes

# Genetic Drift (cont'd)

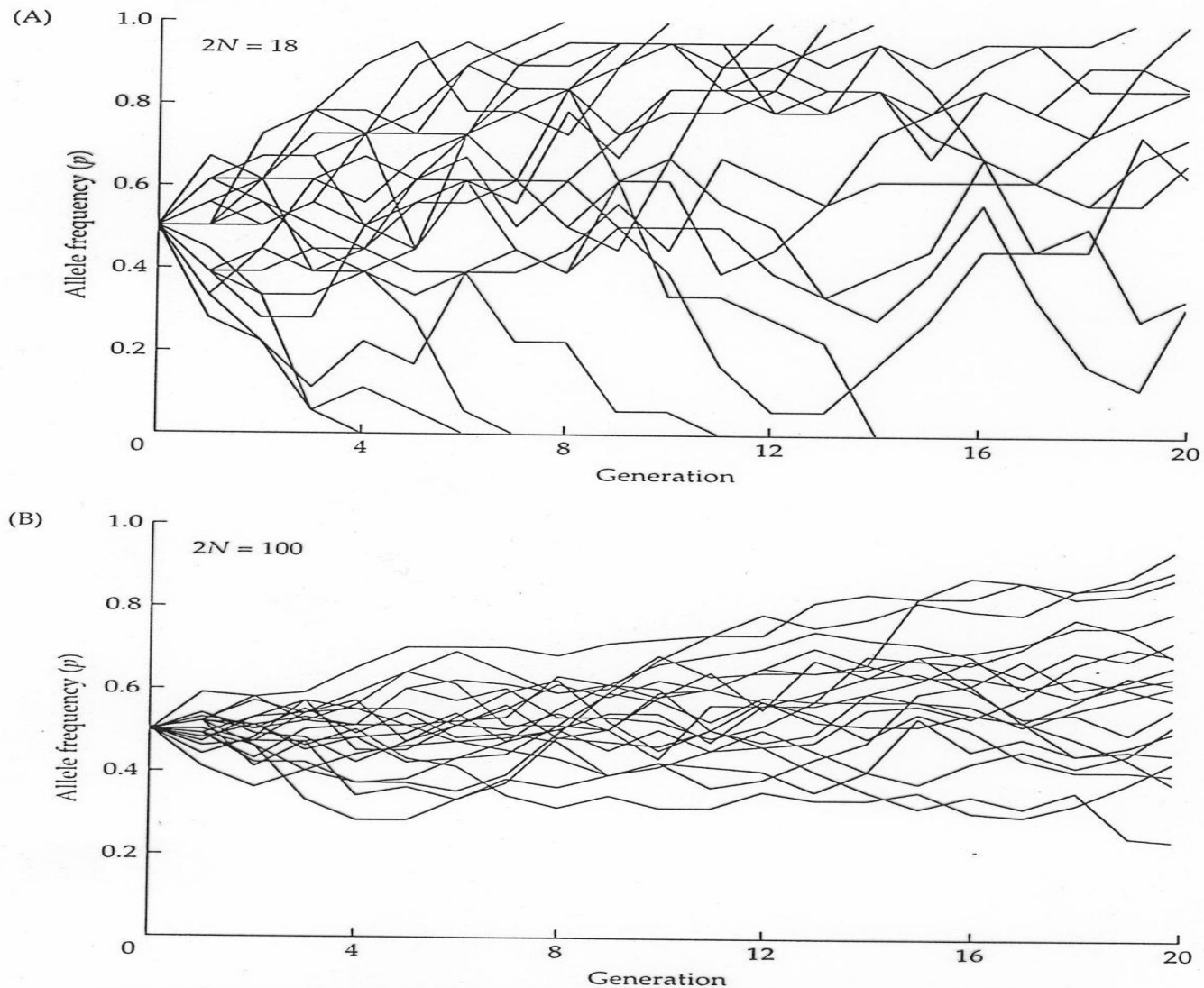
- If # gametes in sample is  $2N$  (where  $N$  = no. of individuals) the probability that the sample contains exactly  $0, 1, 2, 3, \dots, 2N$  alleles of type  $A$  is given by successive terms of the expansion  $(p_A + q_A)^{2N}$

Probability sample contains  $i$  alleles of type  $A$  is  $[(2N)! / i!(2N-i)!] p^i q^{2N-i}$  (binomial probability)



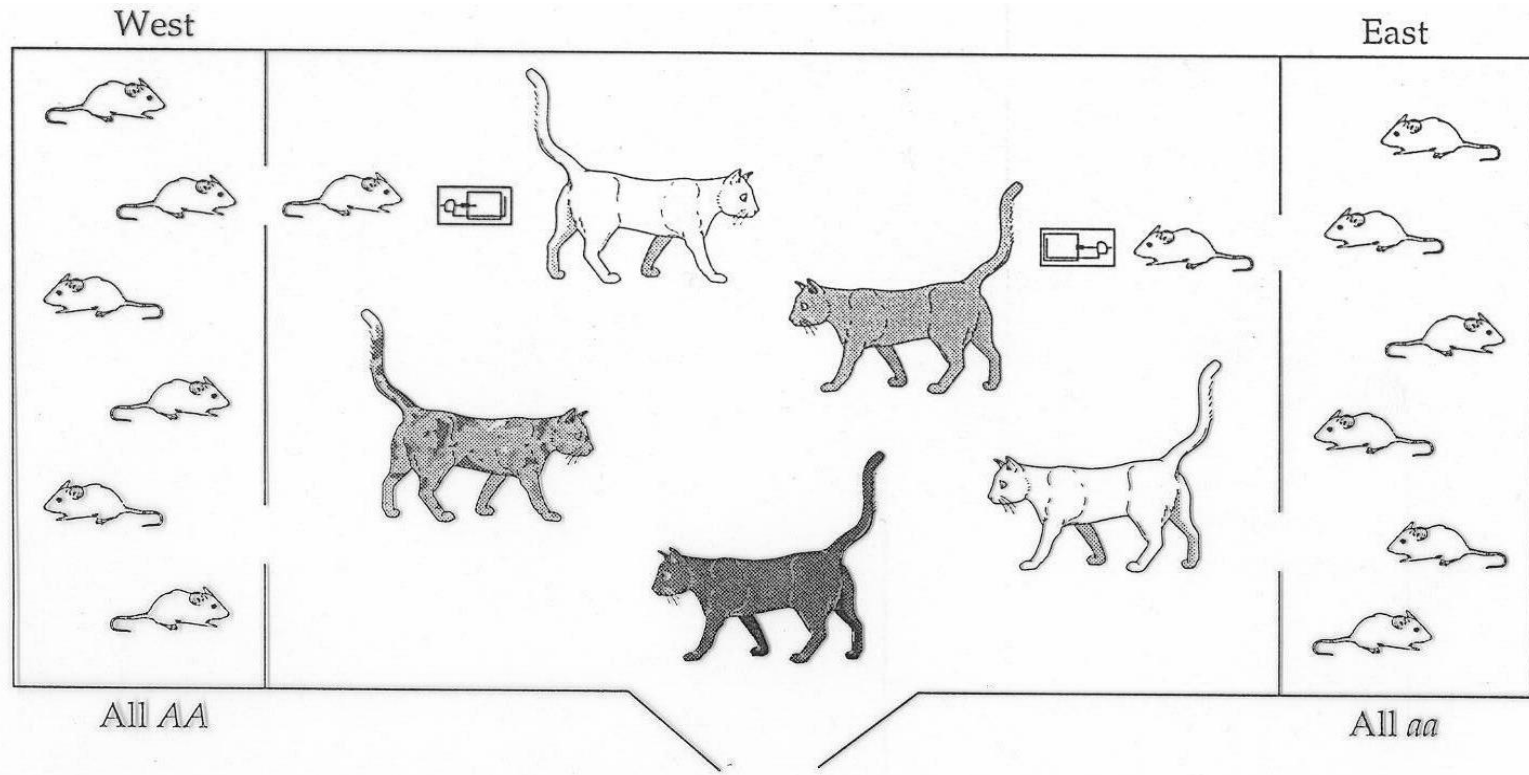
**Figure 2. Model for analyzing effects of random genetic drift. Each of the subpopulations founded from the initial large population (shown by vertical columns of boxes and arrows) is assumed to be genetically isolated from the others. Each subpopulation produces an infinite number of gametes, of which  $2N$  are chosen at random to form the next generation's breeding population. Random genetic drift results from sampling error in this process.**





**Figure 7.2** Computer simulations of the Wright-Fisher model of random genetic drift. Each line represents a population of size (A)  $2N = 18$  or (B)  $2N = 100$ , simulated for 20 generations. Each generation alleles are sampled with replacement as described in the text. An allele frequency of  $p = 0.5$  in A implies that there are nine copies of the  $A$  allele, and nine copies of the  $a$  allele. In B, an allele frequency of 0.5 implies 50 copies of each allele. Note that the larger population size in B results in smaller oscillations of allele frequency, and a slower rate of fixation.





**Figure 4.1** An extreme example of the general principle that a difference in allele frequency among subpopulations results in a deficiency of heterozygotes. The floor plan is that of a hypothetical barn. The mouse subpopulations in the east and west enclaves are completely isolated owing to the cats in the middle. The west subpopulation is fixed for the  $A$  allele and the east subpopulation for the  $a$  allele. Trapping mice at random in the area patrolled by the cats would yield an overall allele frequency of  $1/2$  but no heterozygotes.

# Sub-Population Theory

- Structured population-similar theory to inbreeding

$$A_i A_i: P_{ii} = p_i^2 + p_i (1-p_i)\theta_{ii}$$

$$A_i A_j: P_{ij} = 2p_i p_j (1 - \theta_{ij}) , i \text{ not equal to } j$$

- $\theta_{ij}$  is the probability that 2 alleles drawn at random from the population (not necessarily from the same person) would be ibd.

# NRC REPORT 1996

**Recommendation 4.1:** In general, the calculation of a profile frequency should be made with the product rule. If the race of the person who left the evidence-sample DNA is known, the database for the person's race should be used; if the race is not known, calculations for all racial groups to which possible suspects belong should be made. For systems such as VNTRs, in which a heterozygous locus can be mistaken for a homozygous one, if an upper bound on the genotypic frequency at an apparently homozygous locus (single band) is desired, then twice the allele (bin) frequency,  $2p$ , should be used instead of  $p^2$ . For systems in which exact genotypes can be determined,  $p^2 + p(1-p)\bar{\theta}$  should be used for the frequency at such a locus instead of  $p^2$ . A conservative value of  $\bar{\theta}$  for the US population is 0.01; for some small, isolated populations, a value of 0.03 may be more appropriate. For both kinds of systems,  $2p_i p_j$  should be used for heterozygotes.

A more conservative value of  $\bar{\theta} = 0.03$  might be chosen for PCR-based systems in view of the greater uncertainty of calculations for such systems because of less extensive and less varied population data than for VNTRs.



**We shall never cease  
from exploration  
And the end of all our  
exploring  
Will be to arrive  
where we started  
And know the place  
for the first time.**

**T. S. Eliot**

**Thank you  
for your  
attention!**

