

# Random Genetic Drift

## Effective Population Size

# Random Genetic Drift

- Hardy-Weinberg equilibrium predicts:
  - 1) Allele frequencies remain constant
  - 2) Genotype frequencies predicted by allele frequencies
- HW model assumes infinite population size
- With **finite** population size, allele frequencies change over time due to sampling
- Random genetic drift: stochastic change in allele frequencies due to finite sampling of gametes

# Random Genetic Drift

- Haploid population of size  $N$
- Two alleles:  $A$ ,  $a$
- At generation  $t$ 
  - Frequency of  $A$  is  $p$
- Frequency of  $a$  is  $q = (1-p)$

What is frequency of  $A$  at generation  $t + 1$ ?

Probability( $X$ ) =  
# ways  $X$  can happen / Total Possible Outcomes

1. You roll a die. What is the probability that you roll a 4?

2. You roll a die. What is the probability that you roll an even number?

3. You roll a die twice. What is the probability that you roll a 4 twice?

4. You flip a fair coin. What is the probability you flip a head?



5. You flip a fair coin twice. What is the probability you flip a head twice?

6. You flip a fair coin and it flips heads. What is the probability if you flip it again, it will flip heads?

7. You flip a fair coin twice. What is the probability you flip at least one head?

8. A bag of marbles contains 36 blue marbles and 14 red marbles. If you pull out one marble, what is the probability that it is blue?

9. A bag of marbles contains 36 blue marbles and 14 red marbles. If you pull out one marble, what is the probability that it is NOT blue?

10. A population contains 36 AA individuals and 14 aa individuals. If you pull out genotype at random, what is the probability that it is AA?

11. A randomly mating population contains 36 AA individuals, 50 Aa individuals and 14 aa individuals. These individuals each produce an infinite number of gametes. If you reach into the gamete pool and pull out a single allele, what is the probability that it is A?

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$



# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) =$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) = p$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) = p$
  - $\Pr(0A) =$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) = p$
  - $\Pr(0A) = (1 - p)$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) = p$
  - $\Pr(0A) = (1 - p)$
- Randomly select 2 individuals to be parents

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) =$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) = p$
  - $\Pr(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\Pr(2A) = p^2$



# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) = p^2$
  
  - $\text{Pr}(0A) =$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\Pr(1A) = p$
  - $\Pr(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\Pr(2A) = p^2$
  
  - $\Pr(0A) = (1 - p)^2$

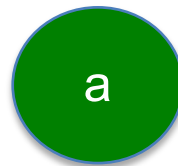
# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) = p^2$
  - $\text{Pr}(1A) =$
  - $\text{Pr}(0A) = (1 - p)^2$



$$\text{Pr} = p(1 - p)$$



$$\text{Pr} = (1 - p)p$$

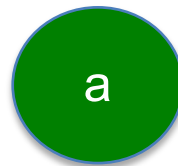
# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) = p^2$
  - $\text{Pr}(1A) = 2p(1-p)$
  - $\text{Pr}(0A) = (1 - p)^2$



$$\text{Pr} = p(1 - p)$$



$$\text{Pr} = (1 - p)p$$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) = p^2$
  - $\text{Pr}(1A) = 2p(1-p)$
  - $\text{Pr}(0A) = (1 - p)^2$
- Randomly select 3 individuals to be parents

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) = p^2$
  - $\text{Pr}(1A) = 2p(1-p)$
  - $\text{Pr}(0A) = (1 - p)^2$
- Randomly select 3 individuals to be parents
  - $\text{Pr}(3A) =$
  - $\text{Pr}(2A) =$
  - $\text{Pr}(1A) =$
  - $\text{Pr}(0A) =$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select 1 individual to be parent
  - $\text{Pr}(1A) = p$
  - $\text{Pr}(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - $\text{Pr}(2A) = p^2$
  - $\text{Pr}(1A) = 2p(1-p)$
  - $\text{Pr}(0A) = (1 - p)^2$
- Randomly select 3 individuals to be parents
  - $\text{Pr}(3A) = p^3$
  - $\text{Pr}(2A) = 3p^2(1 - p)$
  - $\text{Pr}(1A) = 3p(1 - p)^2$
  - $\text{Pr}(0A) = (1 - p)^3$

# Random Genetic Drift

In generation  $t$ ,  $\text{freq}(A) = p$ ,  $\text{freq}(a) = (1-p)$

- Randomly select  $N$  individuals to be parents

$$\Pr(j \text{ alleles of type } A) = \binom{N}{j} p^j (1-p)^{N-j}$$

What is frequency of  $A$  at generation  $t + 1$ ?



# Random Genetic Drift

Frequency of A

Generation(t+1)

0/N

$$\binom{N}{0} (p_t)^0 (1 - p_t)^{N-0}$$

1/N

$$\binom{N}{1} (p_t)^1 (1 - p_t)^{N-1}$$

2/N

$$\binom{N}{2} (p_t)^2 (1 - p_t)^{N-2}$$

·  
·  
·

N/N

$$\binom{N}{N} (p_t)^N (1 - p_t)^0$$

# Random Genetic Drift

Frequency of A	Generation(t+1)	Generation (t+2)
0/N	$\binom{N}{0}(p_t)^0(1-p_t)^{N-0}$	$\binom{N}{0}(p_{t+1})^0(1-p_{t+1})^{N-0}$
1/N	$\binom{N}{1}(p_t)^1(1-p_t)^{N-1}$	$\binom{N}{1}(p_{t+1})^1(1-p_{t+1})^{N-1}$
2/N	$\binom{N}{2}(p_t)^2(1-p_t)^{N-2}$	$\binom{N}{2}(p_{t+1})^2(1-p_{t+1})^{N-2}$
.		
.		
.		
N/N	$\binom{N}{N}(p_t)^N(1-p_t)^0$	$\binom{N}{N}(p_{t+1})^N(1-p_{t+1})^0$

# Random Genetic Drift

Frequency of A	Generation(t+1)	Generation (t+2)	Generation(t+3)
0/N	$\binom{N}{0}(p_t)^0(1-p_t)^{N-0}$	$\binom{N}{0}(p_{t+1})^0(1-p_{t+1})^{N-0}$	$\binom{N}{0}(p_{t+2})^0(1-p_{t+2})^{N-0}$
1/N	$\binom{N}{1}(p_t)^1(1-p_t)^{N-1}$	$\binom{N}{1}(p_{t+1})^1(1-p_{t+1})^{N-1}$	$\binom{N}{1}(p_{t+2})^1(1-p_{t+2})^{N-1}$
2/N	$\binom{N}{2}(p_t)^2(1-p_t)^{N-2}$	$\binom{N}{2}(p_{t+1})^2(1-p_{t+1})^{N-2}$	$\binom{N}{2}(p_{t+2})^2(1-p_{t+2})^{N-2}$
.			
.			
.			
N/N	$\binom{N}{N}(p_t)^N(1-p_t)^0$	$\binom{N}{N}(p_{t+1})^N(1-p_{t+1})^0$	$\binom{N}{N}(p_{t+2})^N(1-p_{t+2})^0$

Transitions between states are random, but defined by a probability

Transitions have no memory beyond previous step

# Random Genetic Drift

Frequency of A	Generation(t+1)	Generation (t+2)	Generation(t+3)
0/N	$\binom{N}{0}(p_t)^0(1-p_t)^{N-0}$	$\binom{N}{0}(p_{t+1})^0(1-p_{t+1})^{N-0}$	$\binom{N}{0}(p_{t+2})^0(1-p_{t+2})^{N-0}$
1/N	$\binom{N}{1}(p_t)^1(1-p_t)^{N-1}$	$\binom{N}{1}(p_{t+1})^1(1-p_{t+1})^{N-1}$	$\binom{N}{1}(p_{t+2})^1(1-p_{t+2})^{N-1}$
2/N	$\binom{N}{2}(p_t)^2(1-p_t)^{N-2}$	$\binom{N}{2}(p_{t+1})^2(1-p_{t+1})^{N-2}$	$\binom{N}{2}(p_{t+2})^2(1-p_{t+2})^{N-2}$
.			
.			
.			
N/N	$\binom{N}{N}(p_t)^N(1-p_t)^0$	$\binom{N}{N}(p_{t+1})^N(1-p_{t+1})^0$	$\binom{N}{N}(p_{t+2})^N(1-p_{t+2})^0$

Transitions between states are random, but defined by a probability

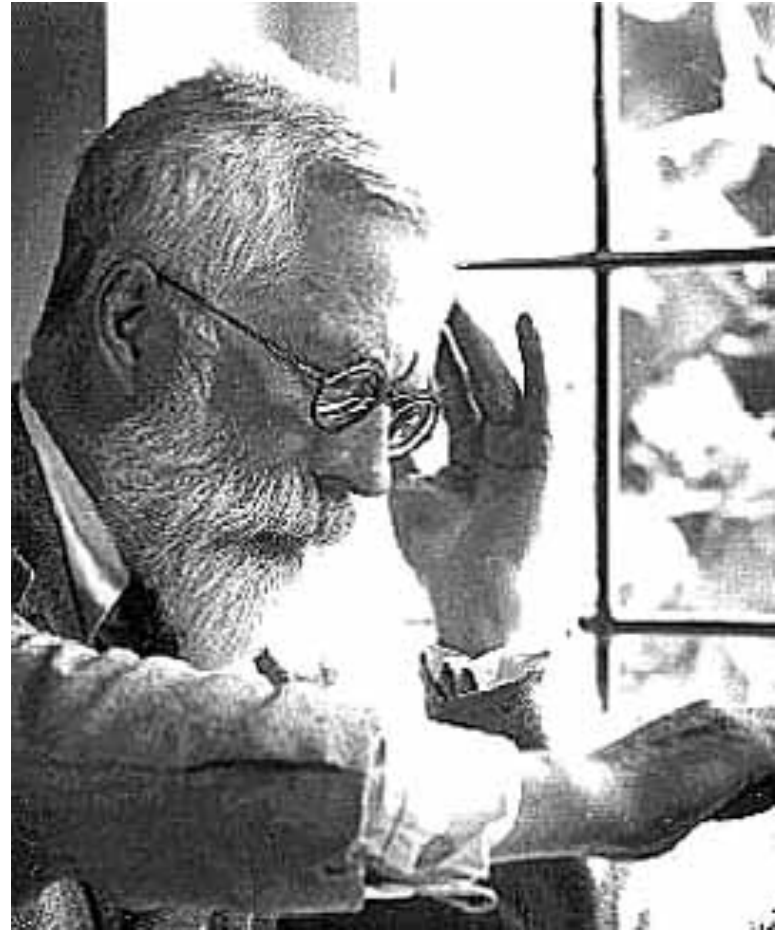
Transitions have no memory beyond previous step

# Diploid Model

- N diploid individuals
  - 2N alleles in population
- Two alleles: A, a (frequencies  $p$ ,  $q$ )
- Randomly draw 2N gametes to found next generation

$$\Pr(j \text{ alleles of type } A) = \binom{2N}{j} p^j (1-p)^{2N-j}$$

# Wright-Fisher Model



# Wright-Fisher Model

- Assumptions:
  - N diploid organisms (2N alleles)
  - Infinite gametes
  - Discrete Generations
  - Random mating
  - No mutation
  - No selection

$$P_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} = \binom{2N}{j} p^j q^{2N-j}$$

# Time, Probability of fixation

- Alleles are eventually fixed or lost
- $2N$  alleles
  - Each equally likely to fix (selectively equivalent)
  - $\text{Pr}(\text{fixation}) =$



# Time, Probability of fixation

- Alleles are eventually fixed or lost
- $2N$  alleles
  - Each equally likely to fix (selectively equivalent)
  - $\text{Pr}(\text{fixation}) = 1/2N$

# Time, Probability of fixation

- Alleles are eventually fixed or lost
- $2N$  alleles
  - Each equally likely to fix (selectively equivalent)
  - $\text{Pr}(\text{fixation}) = 1/2N$
  - If  $i$  copies of allele,  $\text{Pr}(\text{fixation}) = i/2N$

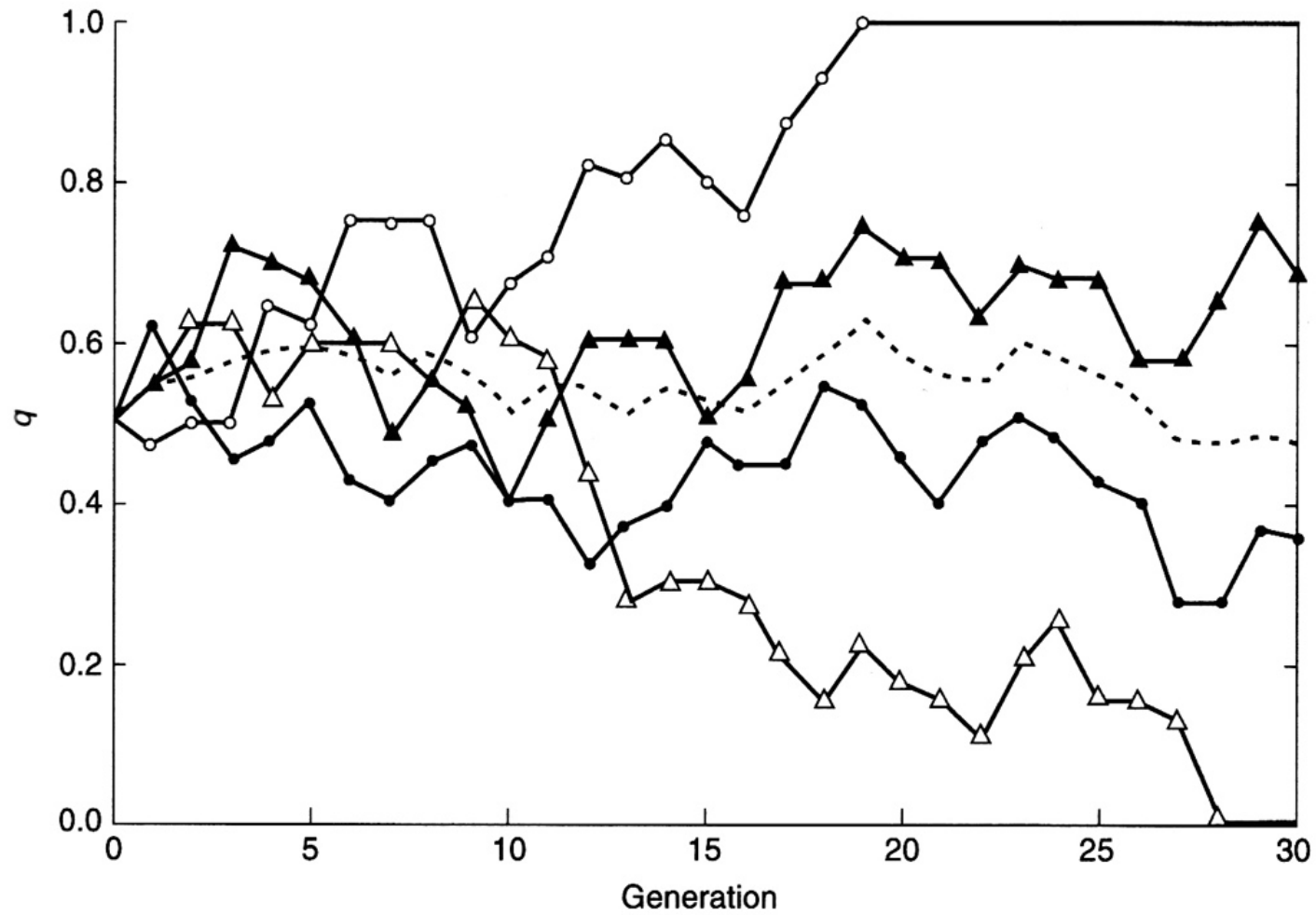
# Time, Probability of fixation

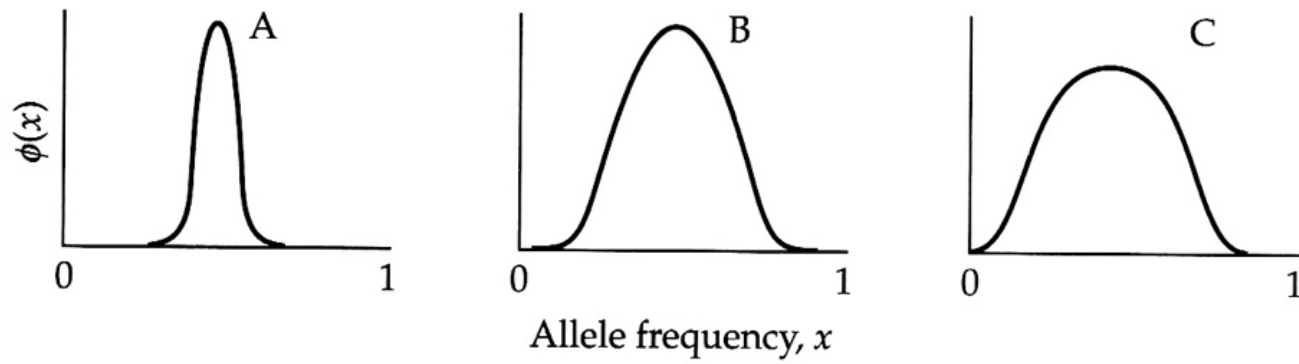
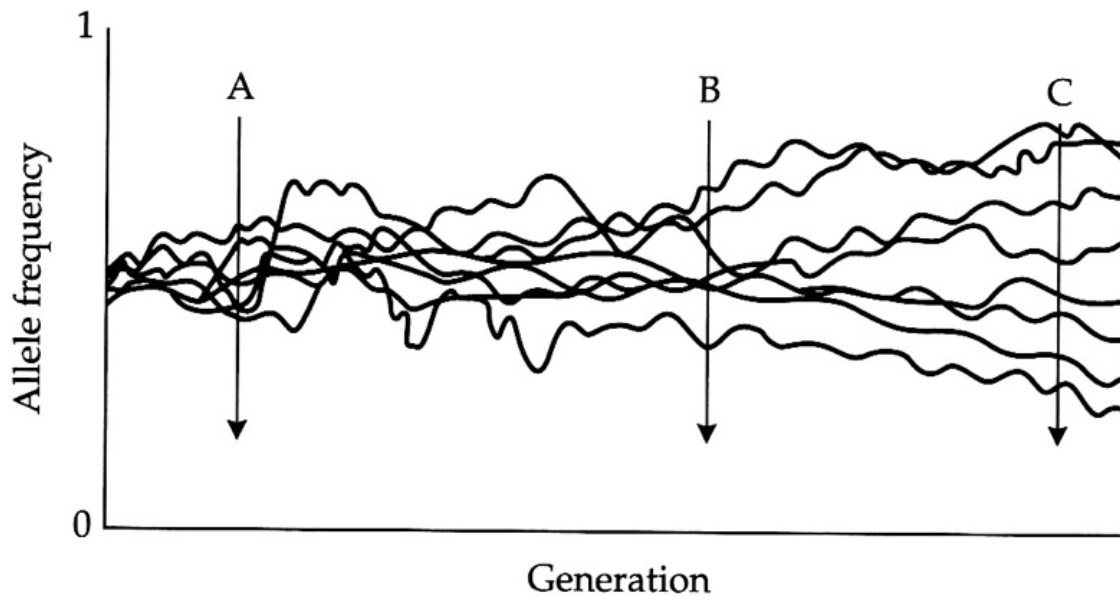
- Alleles are eventually fixed or lost
- $2N$  alleles
  - Each equally likely to fix (selectively equivalent)
  - $\text{Pr}(\text{fixation}) = 1/2N$
  - If  $i$  copies of allele,  $\text{Pr}(\text{fixation}) = i/2N$
- $\text{Pr}(\text{fixation}) = p$

# Probabilities of fixation, loss

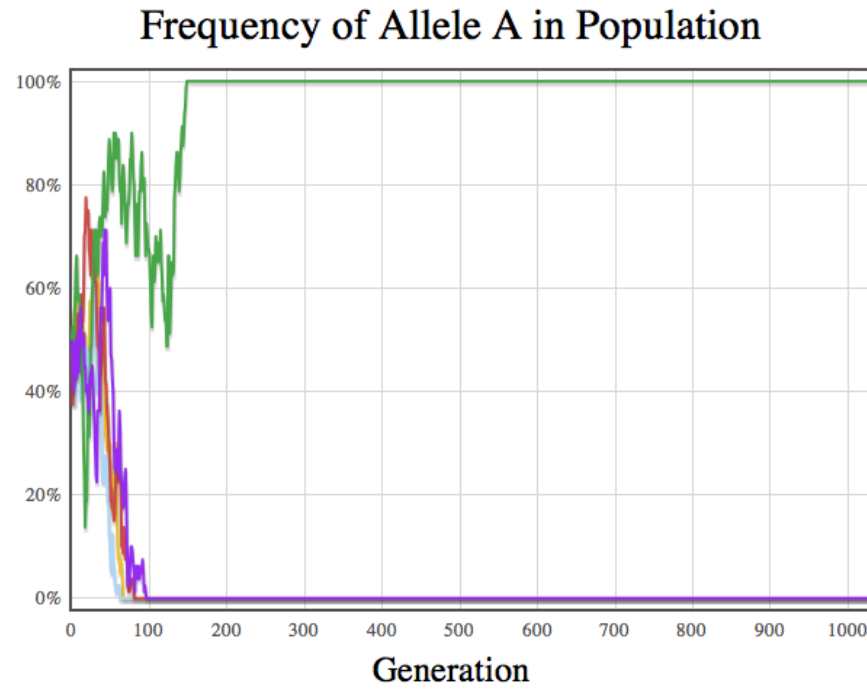
- Alleles are eventually fixed or lost
- $2N$  alleles
  - Each equally likely to fix (selectively equivalent)
  - $\text{Pr}(\text{fixation}) = 1/2N$
  - If  $i$  copies of allele,  $\text{Pr}(\text{fixation}) = i/2N$
- $\text{Pr}(\text{fixation}) = p$
- $\text{Pr}(\text{loss}) = 1-p$

Allele frequencies will change randomly over time

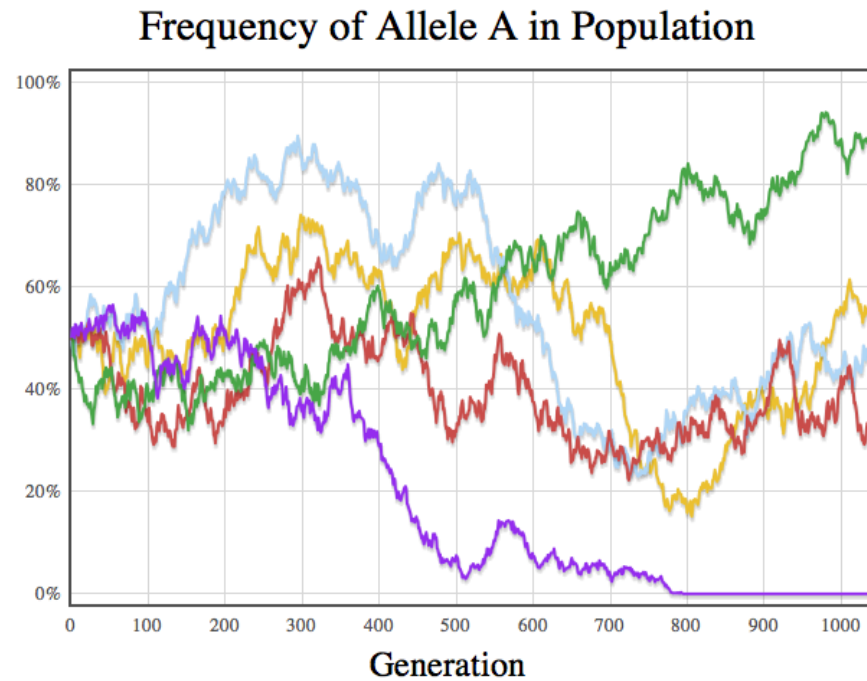




N = 40



N = 2000



# Times to fixation, loss

$$t_{fix} = \frac{-4N(1-p)\ln(1-p)}{p}$$

For  $p = 1/2N$ ,  $t_{fix} \approx 4N$

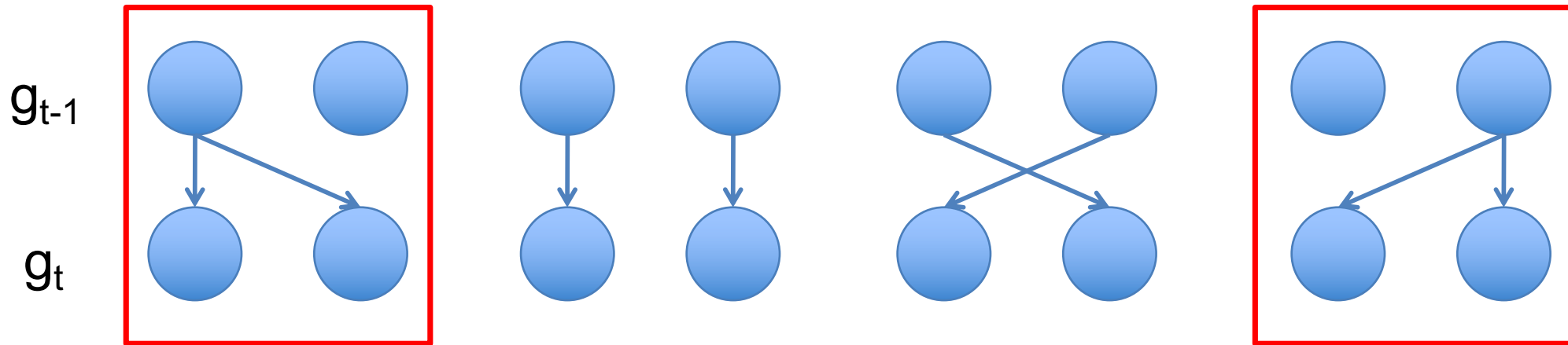
$$t_{loss} = \frac{-4N(p)\ln(p)}{1-p}$$

For  $p = 1/2N$ ,  $t_{loss} \approx 2\ln(2N)$



# Fixation index

$$2N=2$$



$$\Pr(\text{IBD}) = 2/4 = 1/2$$

$$\Pr(\text{IBD}) = 1/2N$$

$$\Pr(\text{not IBD}) = 1 - 1/2N$$

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{t-1}$$

$$\text{If } F_0 = 0, F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

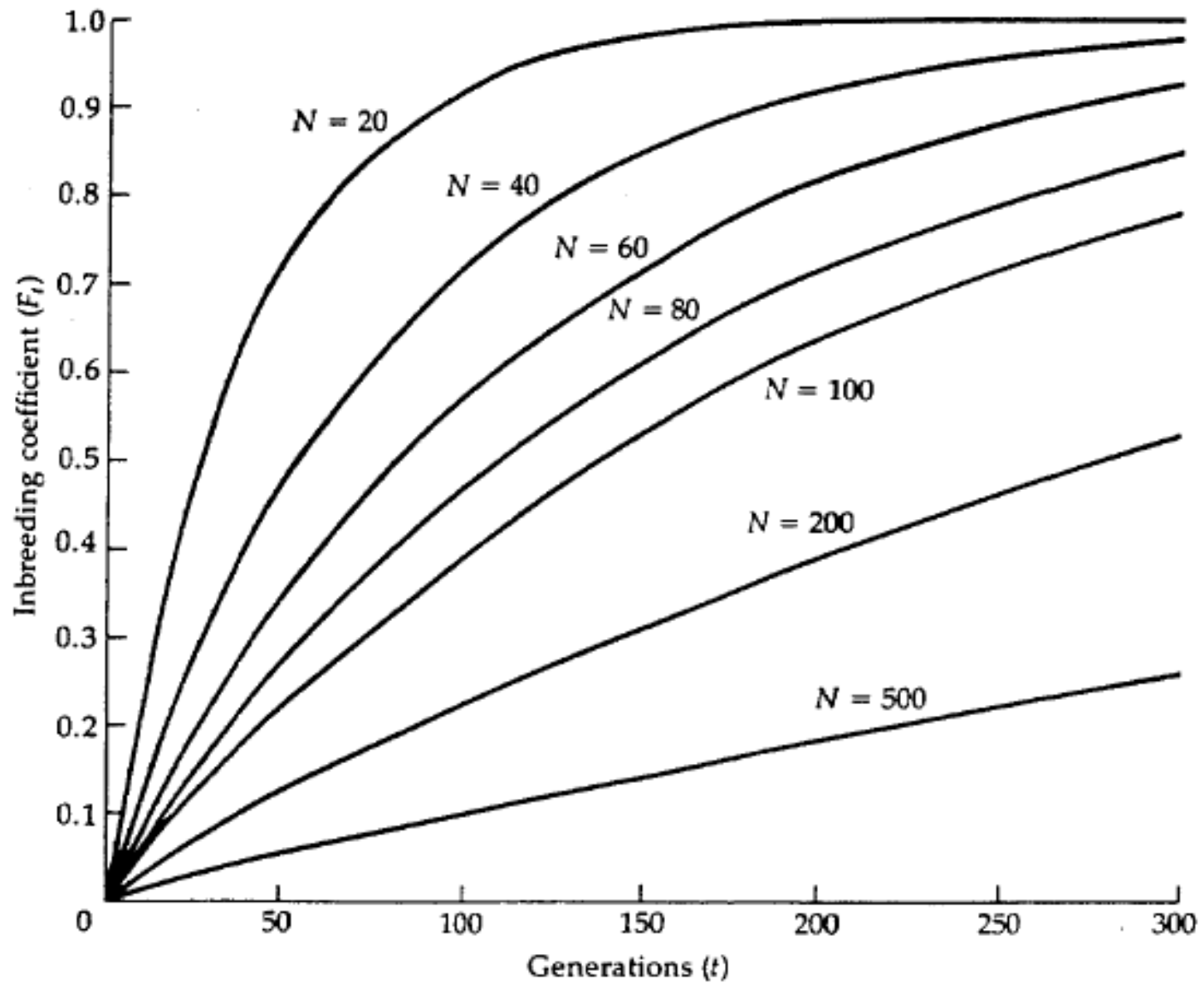


Figure 9. Increase of  $F_t$  in ideal populations as a function of time and effective population size  $N$ .

# Decay of Heterozygosity

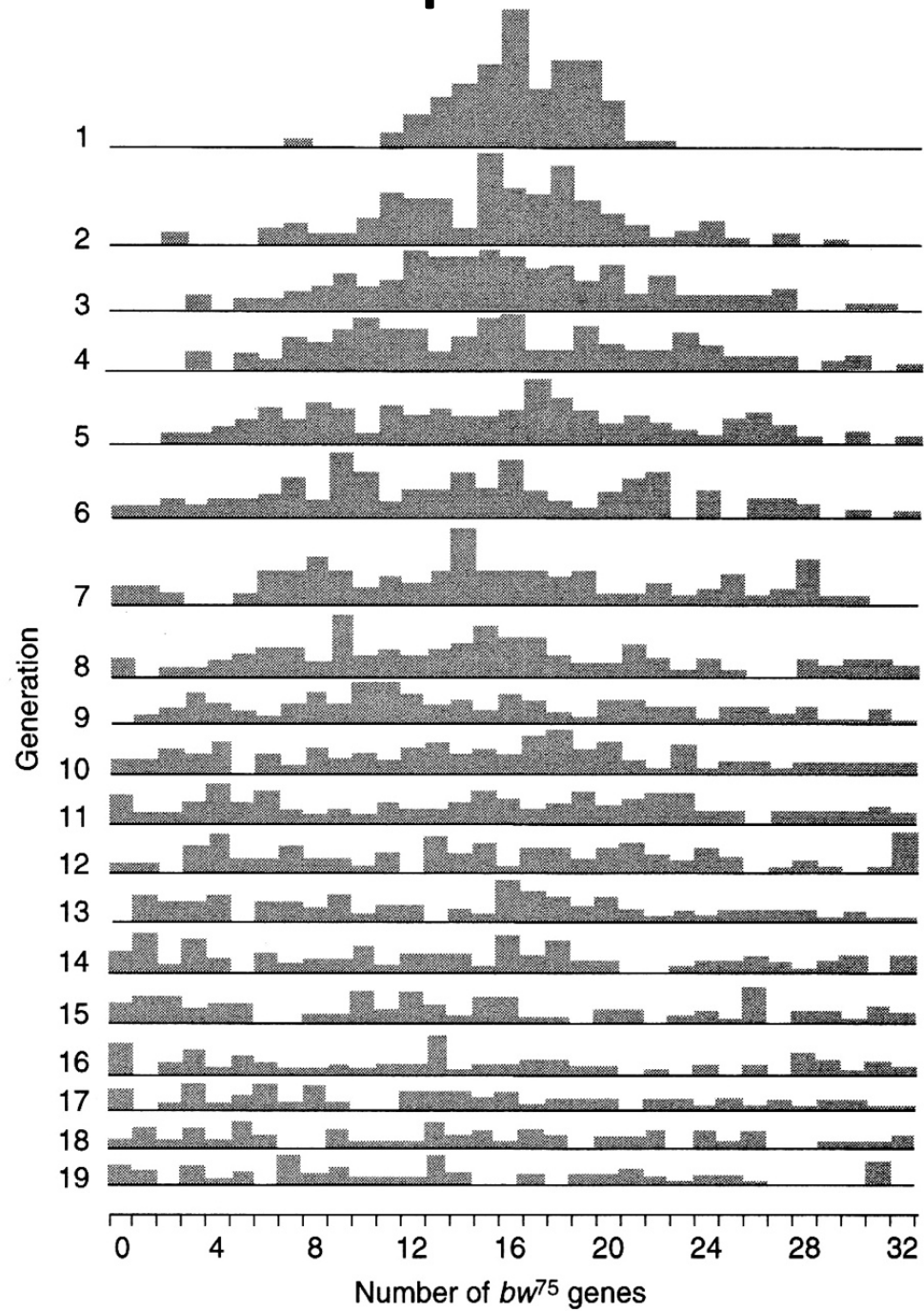
$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

$$H_t = H_0 \left(1 - \frac{1}{2N}\right)^t \approx H_0 e^{-t/2N}$$

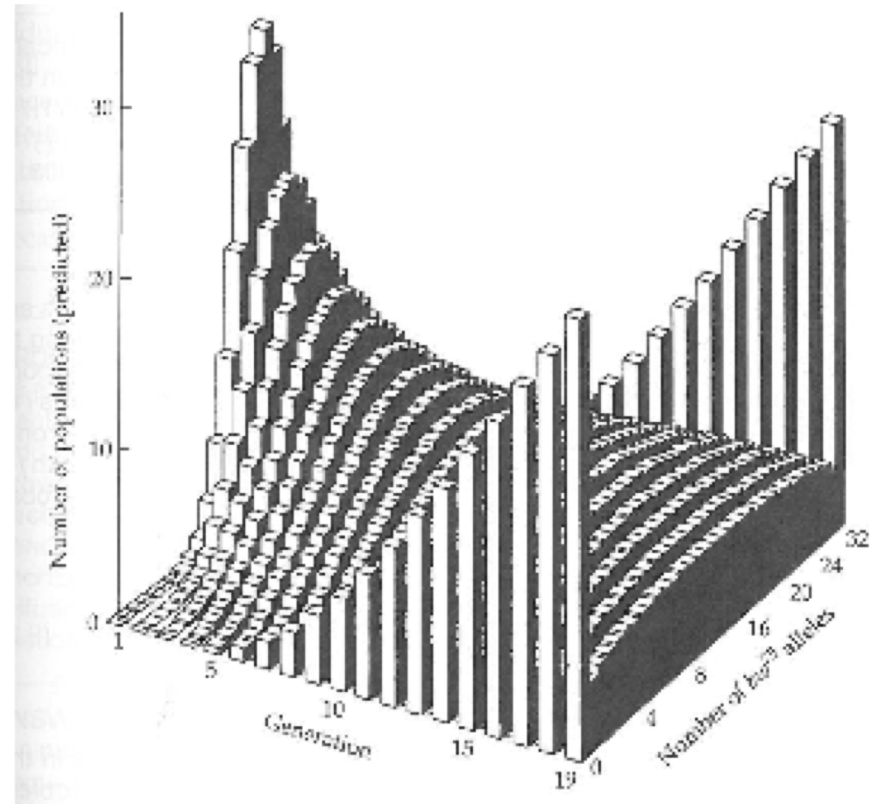
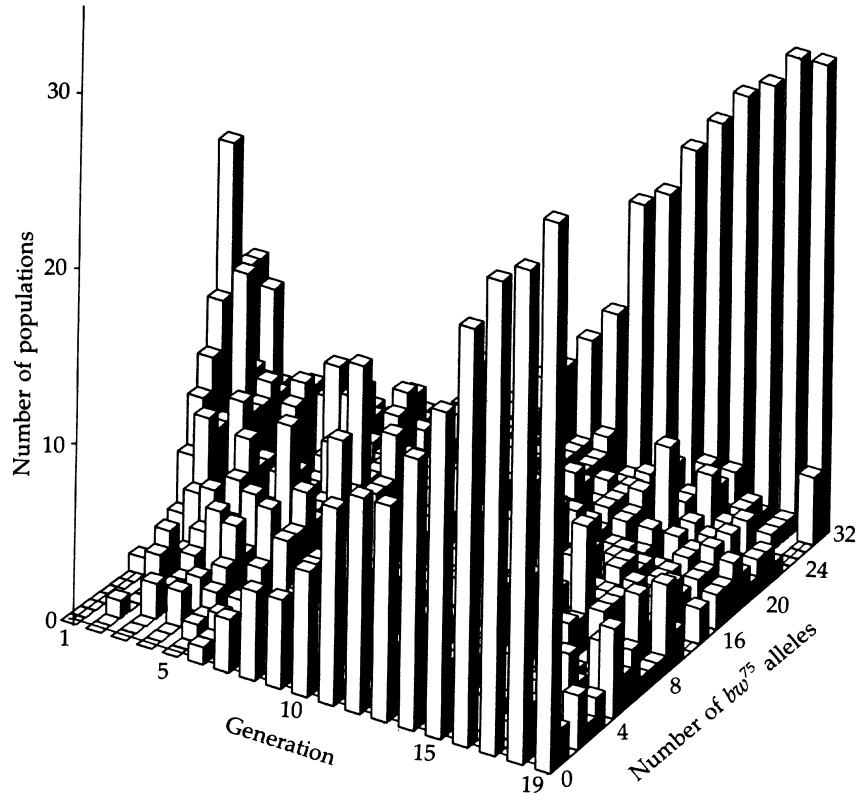
# Summary of Drift

- Stochastic fluctuations in allele frequencies due to sampling in a finite population
- Described by Wright-Fisher model
- Alleles are ultimately fixed or lost from population
  - $\text{Pr}(\text{fix}) = p$ ;  $\text{Pr}(\text{loss}) = 1 - p$
  - $\text{Time}(\text{fix}_{1/2N}) \approx 4N$ ;  $\text{Time}(\text{loss}_{1/2N}) \approx 2\ln(2N)$
- $H$  decreases ( $\sim 1/2N$ ) over time
- Hardy-Weinberg largely still applies
  - Allele frequency changes small
  - Deviation from expected genotype frequencies  $\sim 1/2N$

# Effective Population Size



# Effective Population Size



# Effective Population Size

$$\text{Var}(p) = \text{Var}\left(\frac{\#A}{2N}\right) = \left(\frac{1}{2N}\right)^2 \text{Var}(\#A)$$

$$= \left(\frac{1}{2N}\right)^2 2Npq$$

$$= \frac{pq}{2N}$$

# Effective Population Size

$$\text{Var}(p) = \text{Var}\left(\frac{\#A}{2N}\right) = \left(\frac{1}{2N}\right)^2 \text{Var}(\#A)$$

$$= \left(\frac{1}{2N}\right)^2 2Npq$$

$$= \frac{pq}{2N}$$

- In reality, allele frequency changes  $> pq/2N$ 
  - Fluctuations in population size,  $N_m \neq N_f$ , *etc.*



# Effective Population Size

- Number of individuals in a theoretically ideal population having the same magnitude of drift as the actual population
- Measure 'magnitude' in 3 ways
  - Change in variance of allele frequency (variance  $N_e$ )

$$\text{var}(p) = \frac{pq}{2N}$$

$$\overline{\text{var}}(p) = \frac{pq}{2N_e}$$

$$N_e = \frac{pq}{2\overline{\text{var}}(p)}$$

# Effective Population Size

- Number of individuals in a theoretically ideal population having the same magnitude of drift as the actual population
- Measure 'magnitude' in 3 ways
  - Change in variance of allele frequency (variance  $N_e$ )
  - Change in  $\text{Pr}(\text{IBD})$  (inbreeding  $N_e$ )
  - Rate of loss of heterozygosity (eigenvalue  $N_e$ )

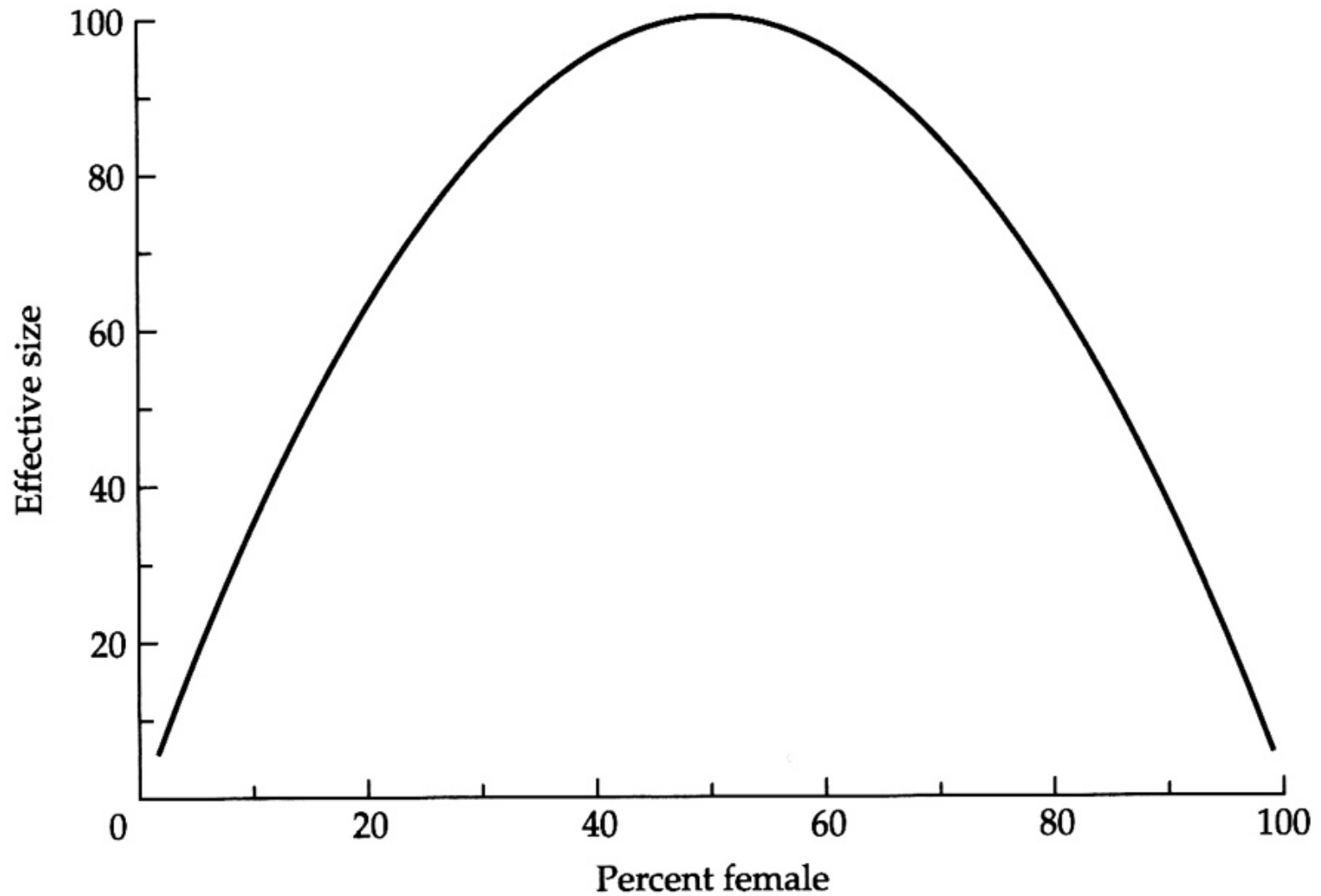
# Effective Population Size

- Number of individuals in a theoretically ideal population having the same magnitude of drift as the actual population
- Measure 'magnitude' in 3 ways
  - Change in variance of allele frequency (variance  $N_e$ )
  - Change in  $\text{Pr}(\text{IBD})$  (inbreeding  $N_e$ )
  - Rate of loss of heterozygosity (eigenvalue  $N_e$ )
- Typically smaller than census size  $N$

# Effective Population Size

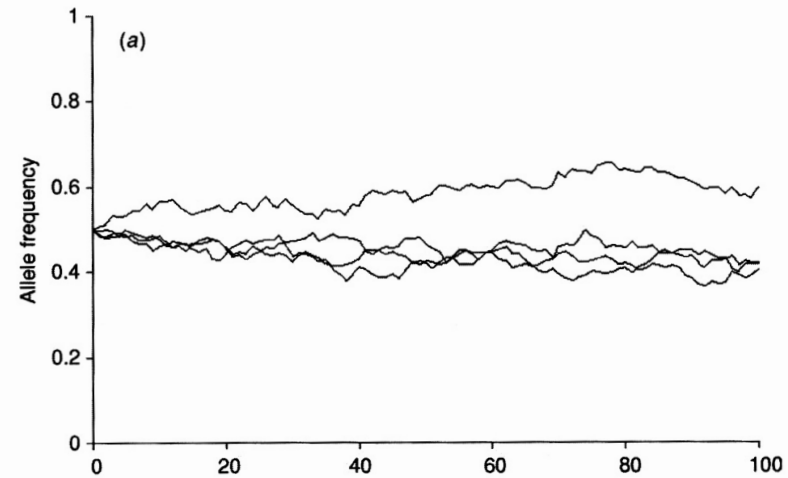
- Number of individuals in a theoretically ideal population having the same magnitude of drift as the actual population
- Measure 'magnitude' in 3 ways
  - Change in variance of allele frequency (variance  $N_e$ )
  - Change in  $\text{Pr}(\text{IBD})$  (inbreeding  $N_e$ )
  - Rate of loss of heterozygosity (eigenvalue  $N_e$ )
- Typically smaller than census size  $N$ 
  - Sex ratio
  - Variance in reproductive success
  - Population size changes

# Effective Population Size

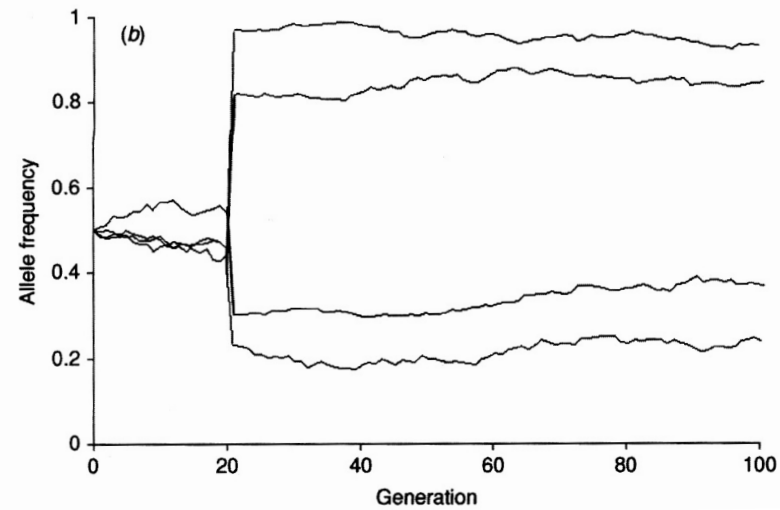


# Effective Population Size

No bottleneck



Bottleneck at  
generation 20



# Relating $N$ , $N_e$

- Population size changes

$$\frac{1}{N_e} = \frac{1}{t} \left( \frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N} \right)$$

# Relating $N$ , $N_e$

- Population size changes

$$\frac{1}{N_e} = \frac{1}{t} \left( \frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N} \right)$$

- Unequal males, females

$$N = N_m + N_f$$

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$