Introductory Models,
Effective Population Size

Models

- Intentional simplification of complex relationships
  - Eliminate extraneous detail, focus on key parameters
  - Appropriate and useful first approximations
- Evaluate fit of data to model
  - Poor fit may implicate violation of model assumptions
  - Refining of models tells us which parameters most important
- Population genetics relies heavily on mathematical models
  - Specify the mathematical relationships among parameters that characterize a population

Random Mating

- One of the most important models in population genetics
- Frequency of mating pairs determined by genotype frequencies

Male Genotype Frequency	Female Genotype Frequency		
	A_1A_1 (P_F)	A_1A_2 (H_F)	A_2A_2 (Q_F)
A_1A_1 (P_M)			
A_1A_2 (H_M)			
A_2A_2 (Q_M)			

Random Mating

- One of the most important models in population genetics
- Frequency of mating pairs determined by genotype frequencies

Male Genotype Frequency	Female Genotype Frequency		
	A_1A_1 (P_F)	A_1A_2 (H_F)	A_2A_2 (Q_F)
A_1A_1 (P_M)	P_MP_F	P_MH_F	P_MQ_F
A_1A_2 (H_M)	H_MP_F	H_MH_F	H_MQ_F
A_2A_2 (Q_M)	Q_MP_F	Q_MH_F	Q_MQ_F

Random Mating

- One of the most important models in population genetics
- Frequency of mating pairs determined by genotype frequencies
- Also called ‘panmictic’ model

Non-overlapping Generations

Birth
Reproduction
Death

Birth
Reproduction
Death

Birth
Reproduction
Death

Generation t -2

Generation t -1

Generation t

Hardy-Weinberg Model

- Both models convenient first approximations for complex populations
- What happens when we combine them?
- What are consequences of random mating in a non-overlapping generation model?

Godfrey Harold Hardy

Wilhelm Weinberg

HW Model Assumptions

- Discrete generations
- Random mating
- Sexual reproduction
- Diploid
- Bi-allelic locus
- Allele frequencies equal in males, females
- Large population size
- No migration
- No mutation
- No selection

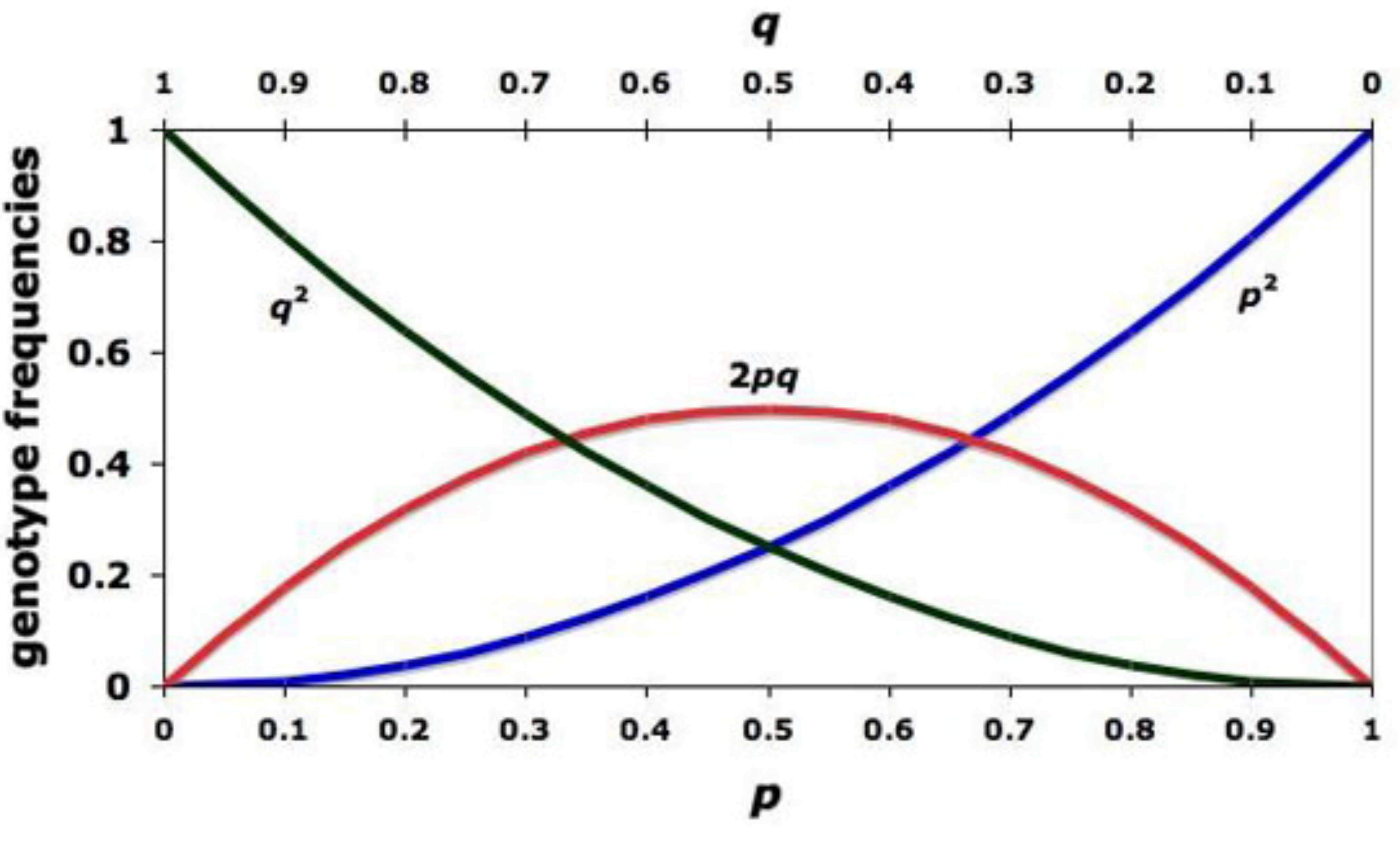
Hardy-Weinberg Principle

- One of first major principles in population genetics
- Describes relationship between genotype frequency and allele frequency
  - Equilibrium state
- Autosomal locus will alleles A, a
  - Frequencies of A, a: p, q
- Genotypes AA, Aa, aa

Hardy-Weinberg Principle

- One of first major principles in population genetics
- Describes relationship between genotype frequency and allele frequency
  - Equilibrium state
- Autosomal locus will alleles A, a
  - Frequencies of A, a: p, q
- Genotypes AA, Aa, aa
  - HW frequencies: p^2, $2pq$, q^2

HWE Genotype Frequencies



Hardy-Weinberg Principle

- One of first major principles in population genetics
- Describes relationship between genotype frequency and allele frequency
  - Equilibrium state
- Autosomal locus will alleles A, a
  - Frequencies of A, a: p, q
- Genotypes AA, Aa, aa
  - HW frequencies: p^2, $2pq$, q^2
- Once at HWE, allele & genotype freq constant

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?

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

- Randomly select 1 individual to be parent

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

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

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

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

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

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

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

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

Random Genetic Drift

In generation t, freq(A) = p, 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, freq(A) = p, 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) =

Random Genetic Drift

In generation t, freq(A) = p, 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, freq(A) = p, 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) =

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

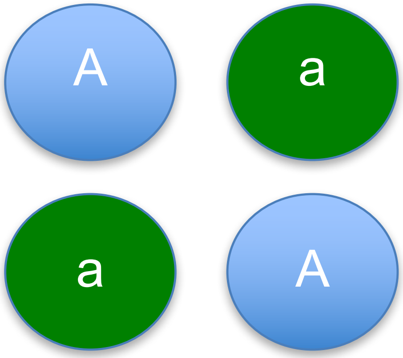
- Randomly select 1 individual to be parent
  - – $\Pr(1A) = p$
  - – $\Pr(0A) = (1 - p)$
- Randomly select 2 individuals to be parents
  - – $\Pr(2A) = p^2$

  - – $\Pr(0A) = (1 - p)^2$

Random Genetic Drift

In generation t, freq(A) = p, 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(1A) =
  - – Pr(0A) = $(1 – p)^2$



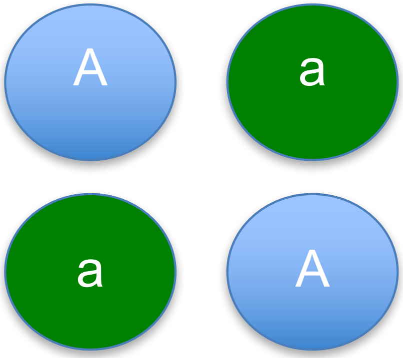
Pr = p(1 – p)

Pr = (1 – p)p

Random Genetic Drift

In generation t, freq(A) = p, 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
  - $\Pr(2A) = p^2$
  - $\Pr(1A) = 2p(1-p)$
  - $\Pr(0A) = (1 - p)^2$



$\Pr = p(1 - p)$

$\Pr = (1 - p)p$

Random Genetic Drift

In generation t, freq(A) = p, 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
  - $\Pr(2A) = p^2$
  - $\Pr(1A) = 2p(1-p)$
  - $\Pr(0A) = (1 - p)^2$
- Randomly select 3 individuals to be parents

Random Genetic Drift

In generation t, freq(A) = p, 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(1A) = $2p(1-p)$
  - Pr(0A) = $(1 - p)^2$
- Randomly select 3 individuals to be parents
  - Pr(3A) =
  - Pr(2A) =
  - Pr(1A) =
  - Pr(0A) =

Random Genetic Drift

In generation t, freq(A) = p, 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
  - – $\Pr(2A) = p^2$
  - – $\Pr(1A) = 2p(1-p)$
  - – $\Pr(0A) = (1 - p)^2$
- Randomly select 3 individuals to be parents
  - – $\Pr(3A) = p^3$
  - – $\Pr(2A) = 3p^2(1 - p)$
  - – $\Pr(1A) = 3p(1 - p)^2$
  - – $\Pr(0A) = (1 - p)^3$

Random Genetic Drift

In generation t, freq(A) = p, freq(a) = (1-p)

- Randomly select N individuals to be parents

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

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

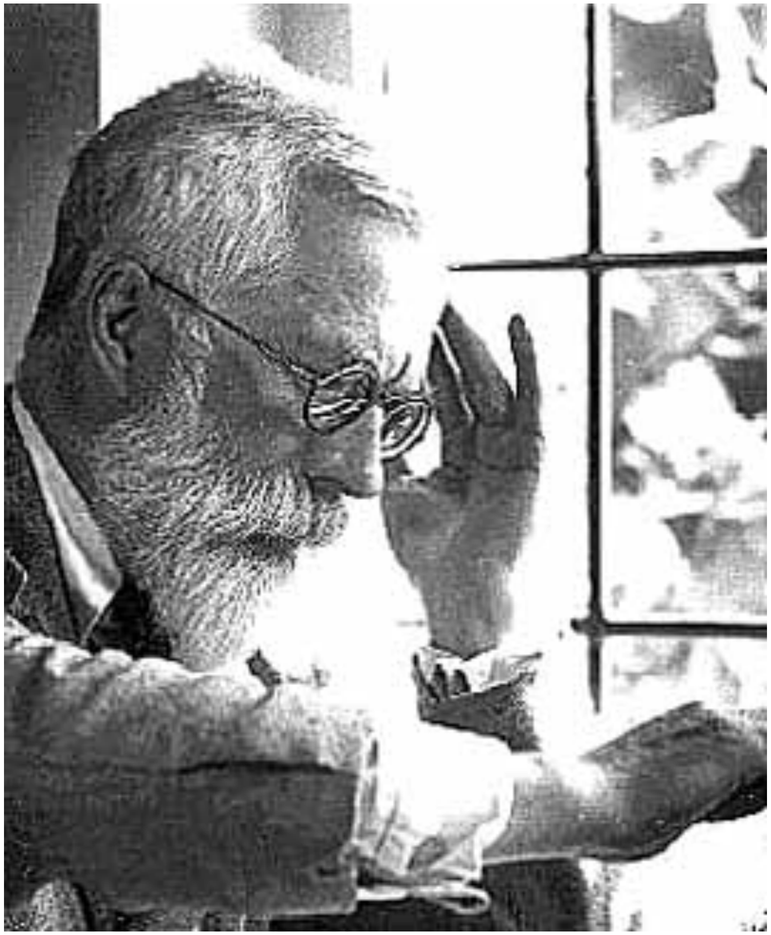
Transitions have no memory beyond previous step

Diploid Model

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

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

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)
  - Pr(fixation) =

Time, Probability of fixation

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

Time, Probability of fixation

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

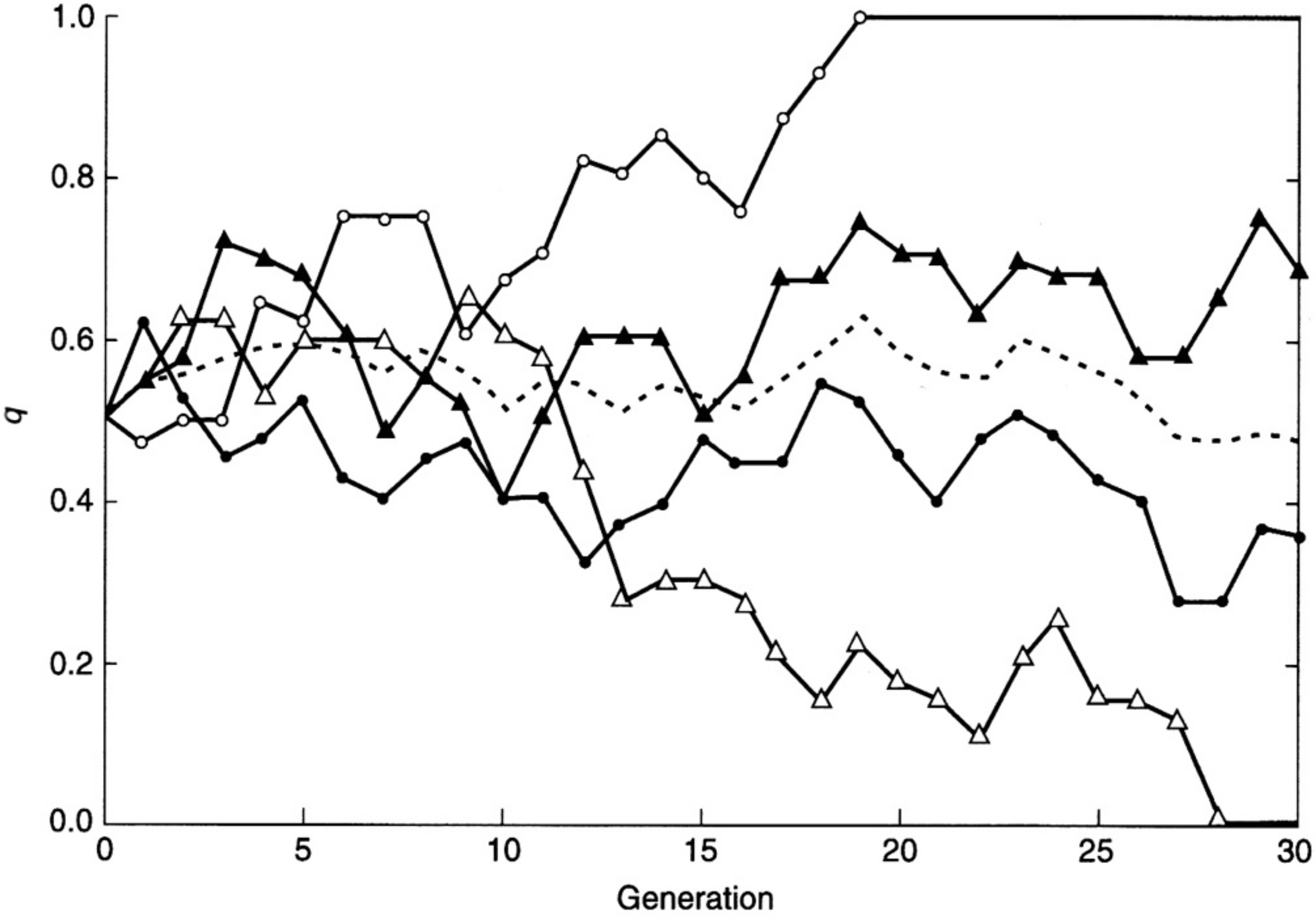
Time, Probability of fixation

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

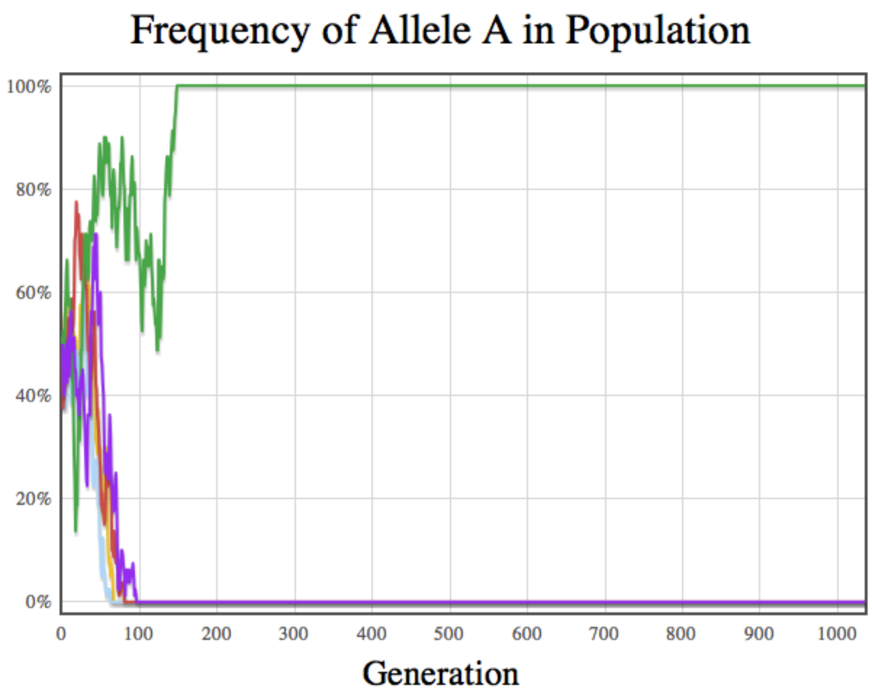
Probabilities of fixation, loss

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

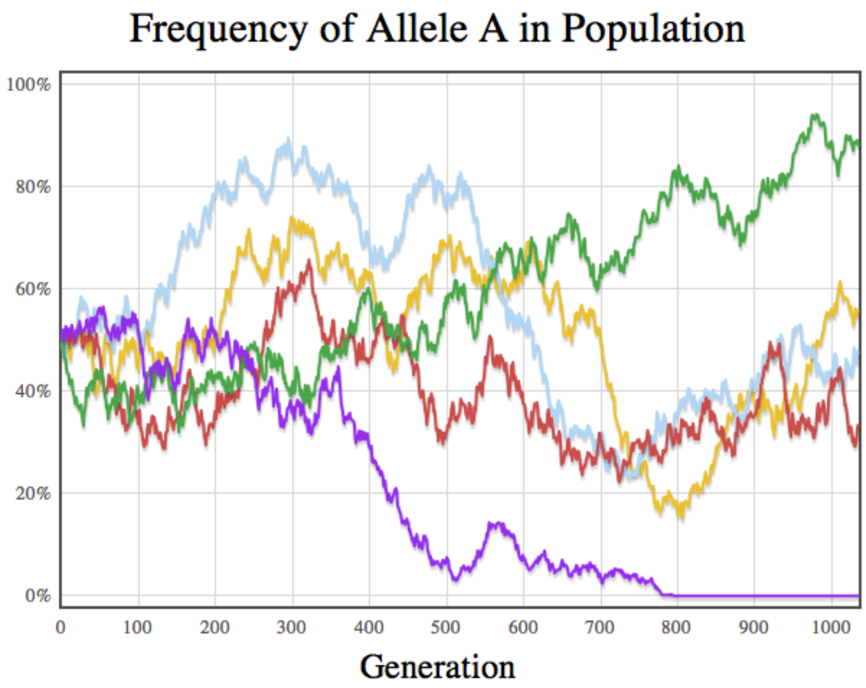
Allele frequencies will change randomly over time



N = 40



N = 2000



http://scit.us/redlynx/

Times to fixation, loss

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

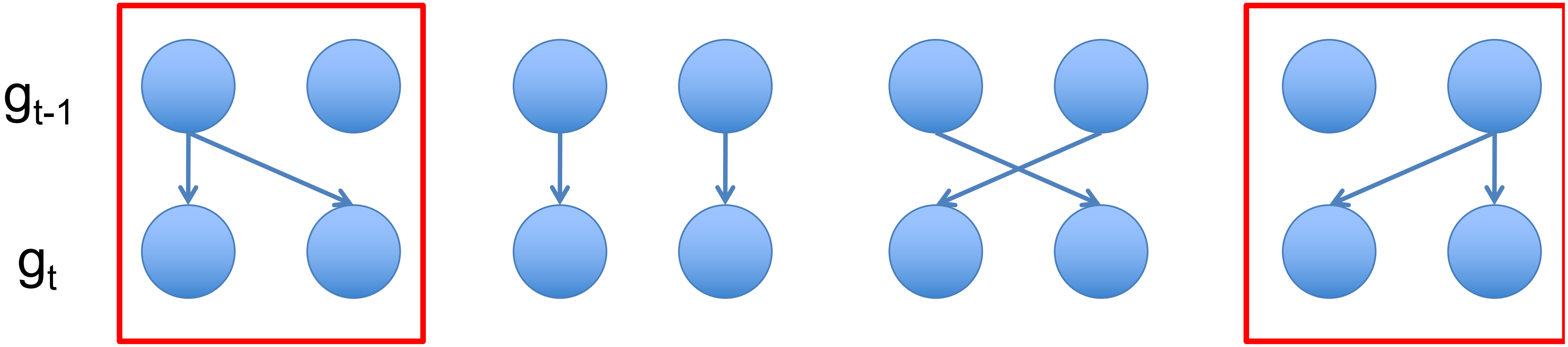
For $p = 1/2N$, $t_{fix} \approx 4N$

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

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

Decay of Heterozygosity

$2N=2$



Pr(IBD) = 2/4 = 1/2

Pr(IBD) = 1/2N

Pr(not IBD) = 1 - 1/2N

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

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

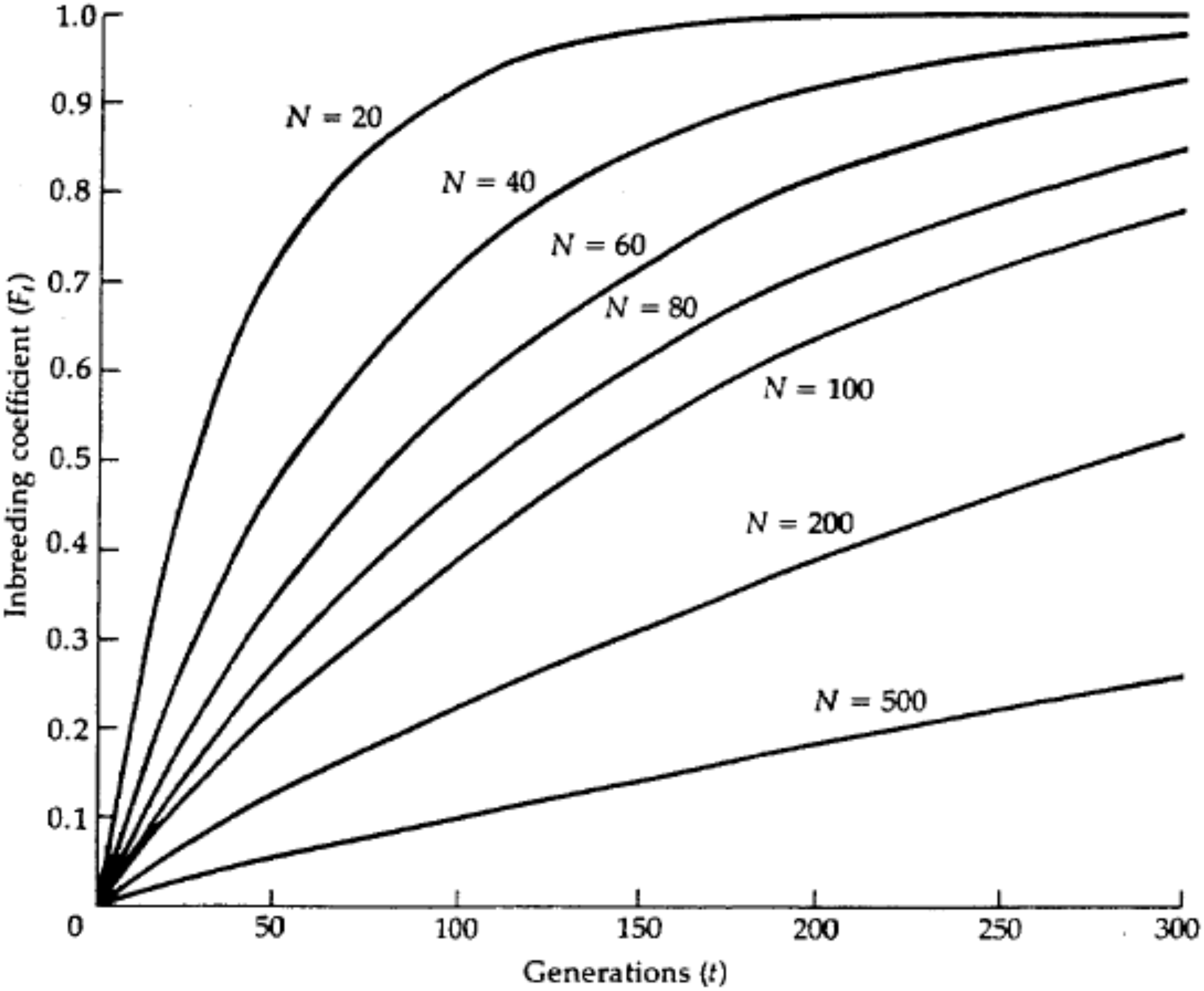


Figure 9. Increase of F_t in ideal populations as a function of time and effective population size N.

Hartl & Clark

Decay of Heterozygosity

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

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

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
  - Pr(fix) = p; Pr(loss) = 1 – p
  - Time($\text{fix}_{1/2N}$) ≈ 4N; Time($\text{loss}_{1/2N}$) ≈ 2ln(2N)
- H decreases (~1/2N) over time
- Hardy-Weinberg largely still applies
  - Allele frequency changes small
  - Deviation from expected genotype frequencies ~ 1/2N

How does drift come into play in conservation genetics?

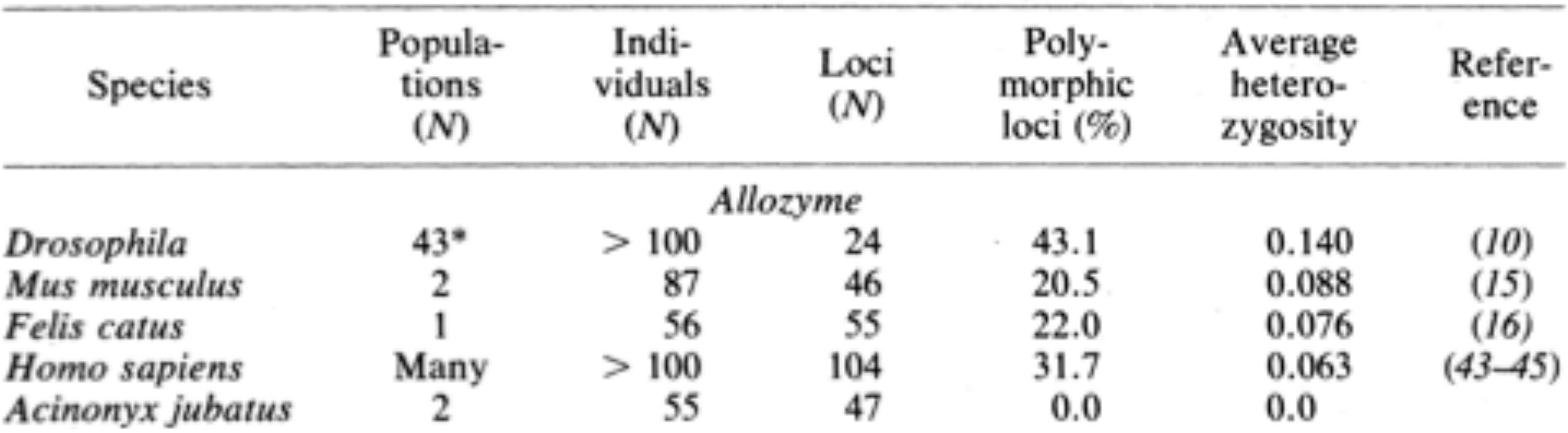
Habitat loss

- Small, isolated populations are threatened
  - RGD

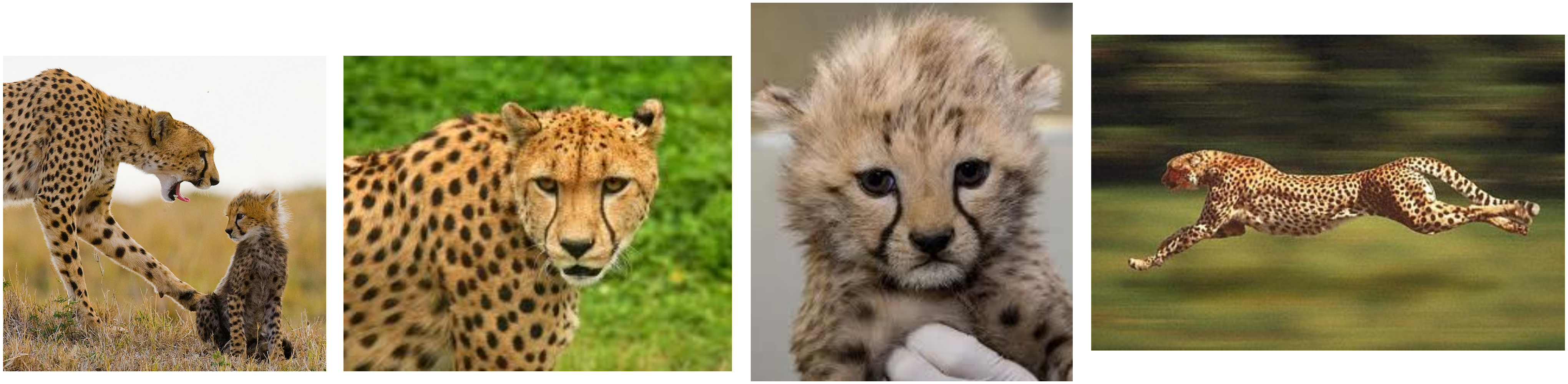
Habitat loss

- Small, isolated populations are threatened
  - RGD, inbreeding
- Consequences:
  - Reduction in genetic diversity

Genetically depauperate species

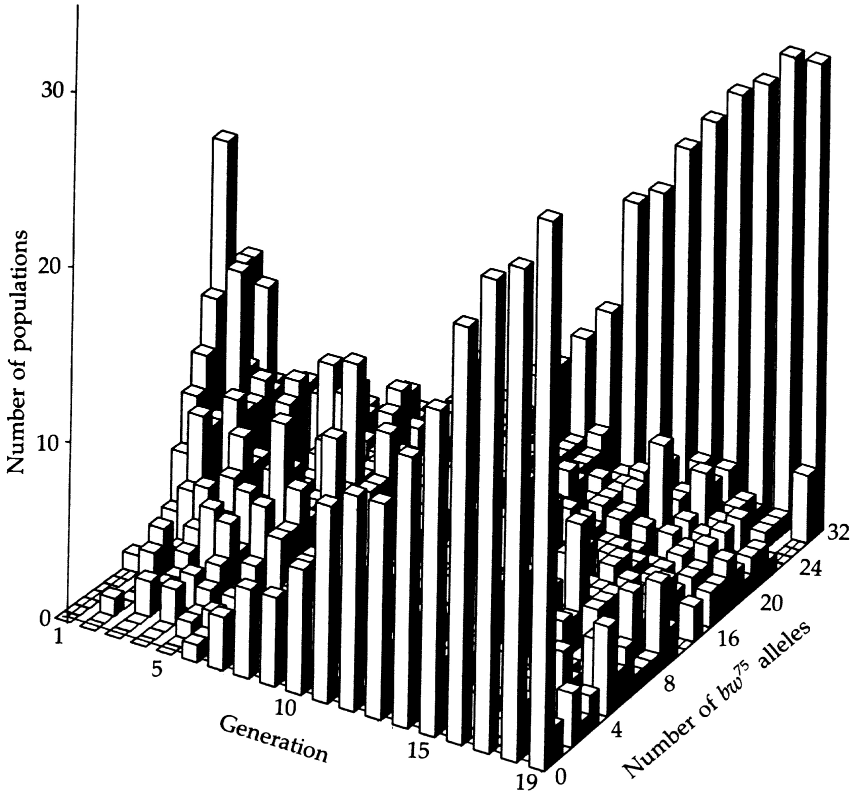


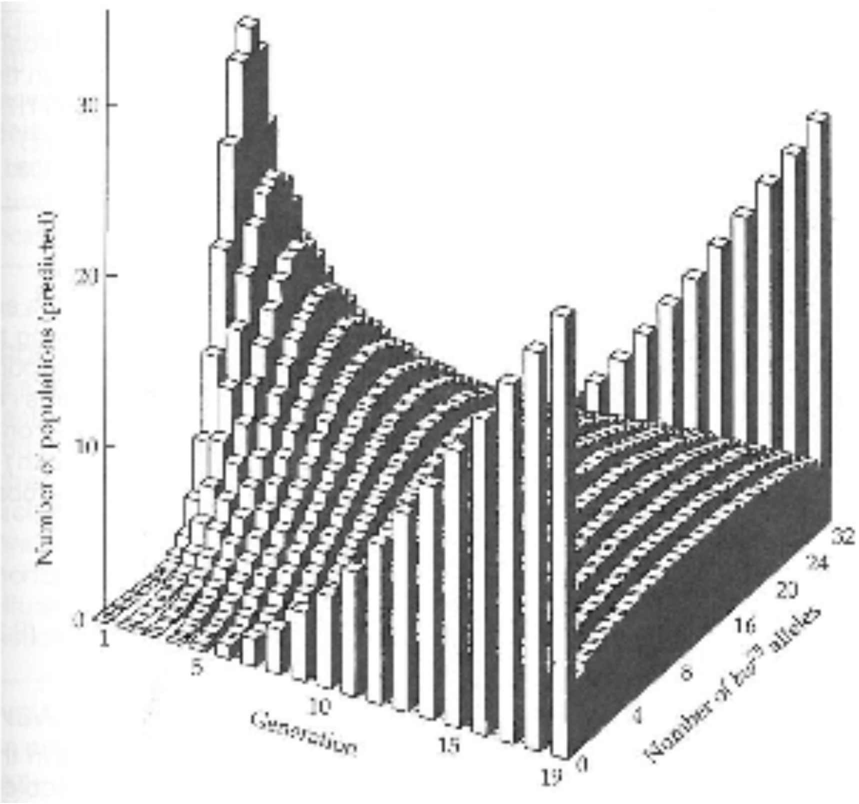
Species	Populations (*N*)	Individuals (*N*)	Loci (*N*)	Polymorphic loci (%)	Average heterozygosity	Reference
			Allozyme			
Drosophila	43*	> 100	24	43.1	0.140	*(10)*
Mus musculus	2	87	46	20.5	0.088	*(15)*
Felis catus	1	56	55	22.0	0.076	*(16)*
Homo sapiens	Many	> 100	104	31.7	0.063	*(43–45)*
Acinonyx jubatus	2	55	47	0.0	0.0	



O’Brien et al. 1983

Effective Population Size





Buri 1956

Effective Population Size

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

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

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

Effective Population Size

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

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

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

- 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

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 Pr(IBD) (inbreeding N_e)
  - Rate of loss of heterozygosity (eigenvalue N_e)

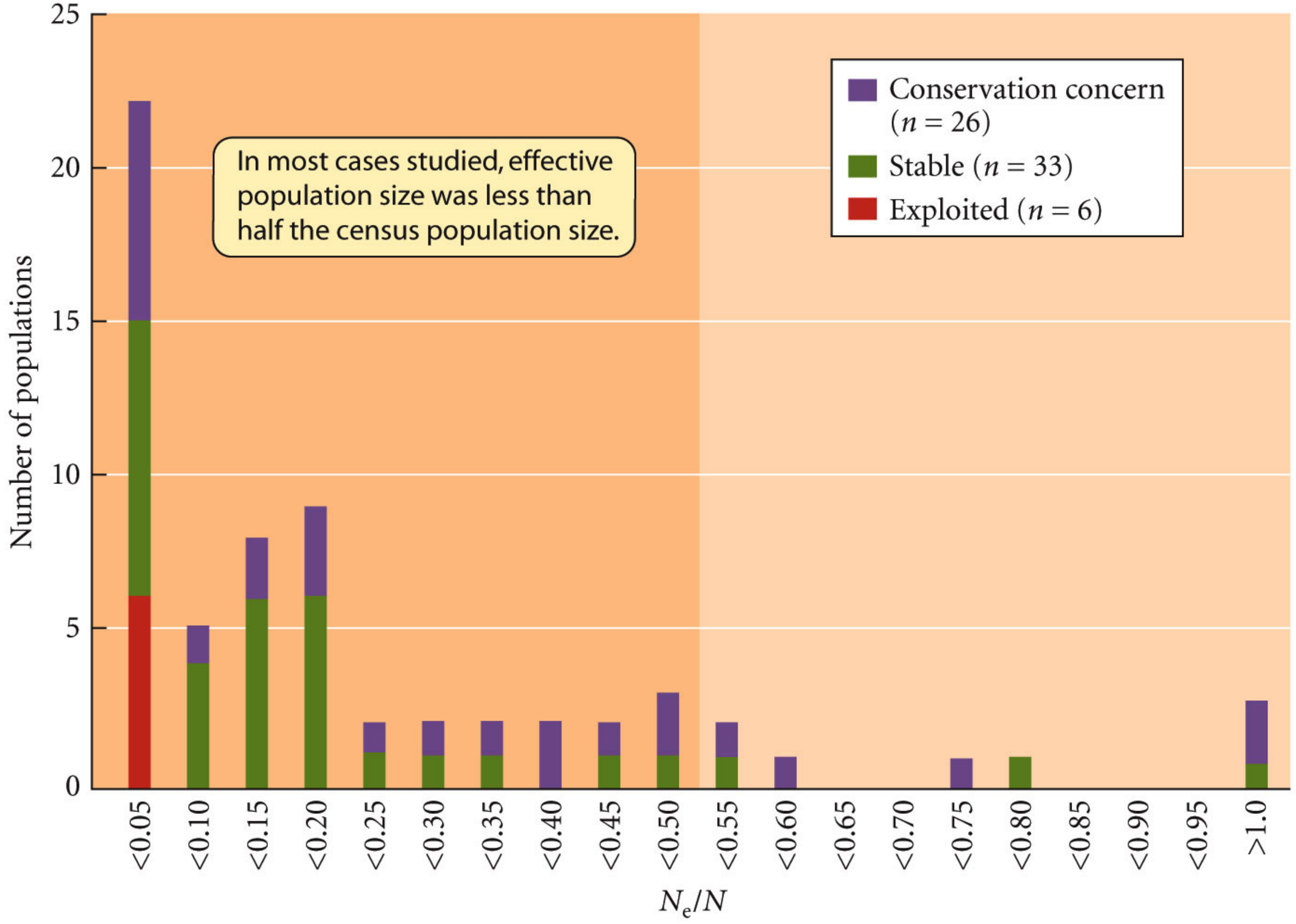
Effective Population Size

- 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 Pr(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 Pr(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



EVOLUTION 2e, Figure 10.6