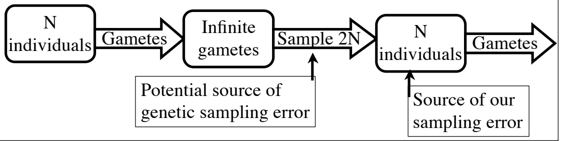
Random Genetic Drift

Why Genetic Variation Matters

Wright-Fisher

Wright-Fisher Breeding

- Adults produce an infinite pool of gametes
  - No sampling error in producing gametes
- Enough gametes are drawn to produce the next generation (2N)
- This produces a binomial distribution gamete frequencies


N individuals

Gametes

Infinite gametes

Sample 2N

N individuals

Gametes

Potential source of genetic sampling error

Source of our sampling error

Genetic Drift

- The breeding system allows for change in frequencies between generations because sampling is limited.
- Often we are only concerned with one allele at a time so we can use the binomial in most cases: 1 allele versus the rest
- The probability that a generation (sample) contains exactly i copies of an allele is given by the Binomial Probability
- The amount of variance we see is inversely proportional to how many times we draw from the gamete pool

Binomial probability

$$P(i) = \binom{2N}{i} p^i q^{2N-i}$$

$$\binom{2N}{i} = \frac{(2N)!}{i!(2N-i)!}$$

Binomial Variance

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

Multinomial probability

$$P = \frac{N!}{\prod_{i-1}^{n} x_i!} \prod_{i=1}^{n} \theta_i^{x_i}$$

$$\prod_{i=1}^{n} \theta_i^{x_i} = p^{x_i} q^{x_j} \ldots$$

$$\sum_{i=1}^{n} x_i = N, \qquad \sum_{i=1}^{n} \theta_i = 1,$$

Wright-Fisher Transitions

- We can predict the probability of changing the number of alleles between generations.
- Example a diploid population with N=10
- The probability is greatest that the allele frequency will change (1.0 - 0.285 = 0.715)
- But the probability is also greatest that the allele frequency is going to stay the same or change only very little (0.285+0.190+0.270 = 0.745)

$$T_{ij} = \binom{2N}{j} p^j q^{2N-j}$$

i = # A alleles time 1
j = # A alleles time 2

$$\binom{2N}{j} = \frac{(2N)!}{j!(2N-j)!}$$

Example n = 10

$$T_{2,2} = \frac{(2*10)!}{2!(2*10-2)!}(0.1)^2 * (0.9)^{(2*10-2)}$$

$$T_{2,2} = \frac{20*19}{2*1}(0.01)*(0.15) = 0.285$$

$$T_{2,3} = \frac{(2*10)!}{3!(2*10-3)!}(0.1)^3 * (0.9)^{(2*10-3)}$$

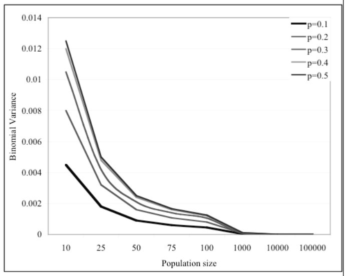
$$T_{2,3} = \frac{20*19*18}{3*2*1}(0.001)*(0.167) = 0.190$$

$$T_{2,1} = \frac{(2*10)!}{1!(2*10-1)!}(0.1)^1 * (0.9)^{(2*10-1)}$$

$$T_{2,1} = \frac{20}{1}(0.1)*(0.135) = 0.270$$

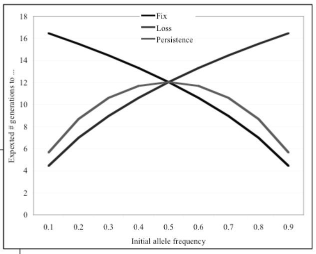
Wright-Fisher Transitions

- Binomial Variance = pq/2N
- The degree of change expected in the population
  - Inversely dependent upon the size of the population
  - Inversely related to the difference between the two allele frequencies



Persistence Time

- Most alleles are doomed the only question is how long will they persist.
- The diffusion equation allows us to estimate how long an allele has until it is food for worms.



$$\bar{t}_{fix} = -\frac{4N}{p}\left[(1-p)\log(1-p)\right]$$

$$\bar{t}_{loss} = -\frac{4N}{1-p}\left[p\log(p)\right]$$

$$\bar{t}_{persistence} = -4N\left[p\log(p) + (1-p)\log(1-p)\right]$$

Persistence Time

- Example
  - N = 10, p = 0.2

$$\bar{t}_{fix} = -\frac{4N}{p}\left[(1-p)\log(1-p)\right]$$

$$\bar{t}_{loss} = -\frac{4N}{1-p}\left[p\log(p)\right]$$

$$\bar{t}_{persistence} = -4N\left[p\log(p)+(1-p)\log(1-p)\right]$$

$$\bar{t}_{fix} = -\frac{4*10}{0.2}\left[(1-0.2)\log(1-0.2)\right] = 15.5$$

$$\bar{t}_{loss} = -\frac{4*10}{0.2}\left[0.2\log(0.2)\right] = 27.96$$

$$\bar{t}_{persistence} = -40\left[0.2\log(0.2)+(1-0.2)\log(1-0.2)\right] = 8.69$$

Genetic Drift vs. Inbreeding

- Drift vs. Inbreeding
  - Decrease H
  - Increase F (autozygosity)
- Assumptions
  - Diploid
  - Sexual reproduction
  - Nonoverlapping generations
  - Many independent subpopulations
  - N constant in subpopulations
  - Random mating
  - No migration
  - No mutation
  - No selection

T = 0	T = n-i	T = n
P=0.5		
P=0.5		
P=0.5		
P=0.5		

Genetic Drift vs. Inbreeding

- Drift vs. Inbreeding
  - Decrease H
  - Increase F (autozygosity)
- Assumptions
  - Diploid
  - Sexual reproduction
  - Nonoverlapping generations
  - Many independent subpopulations
  - N constant in subpopulations
  - Random mating
  - No migration
  - No mutation
  - No selection

T = 0	T = n-i	T = n
P=0.5	P=0.1	
P=0.5	P=0.8	
P=0.5	P=0.9	
P=0.5	P=0.2	

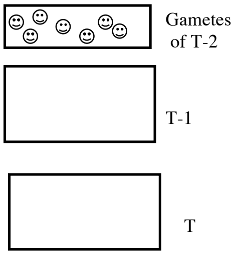
Genetic Drift vs. Inbreeding

- Drift vs. Inbreeding
  - Decrease H
  - Increase F (autozygosity)
- Assumptions
  - Diploid
  - Sexual reproduction
  - Nonoverlapping generations
  - Many independent subpopulations
  - N constant in subpopulations
  - Random mating
  - No migration
  - No mutation
  - No selection

T = 0	T = n-i	T = n
P=0.5	P=0.1	P=0.0
P=0.5	P=0.8	P=1.0
P=0.5	P=0.9	P=1.0
P=0.5	P=0.2	P=0.0

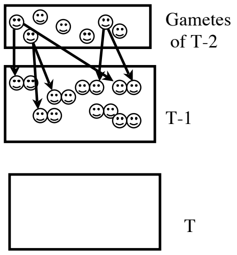
Genetic Drift vs. Inbreeding

- Drift autozygosity
  - Limited population size and random mating
    - $F_t \neq 0$
    - Some individuals will be autozygous by chance.



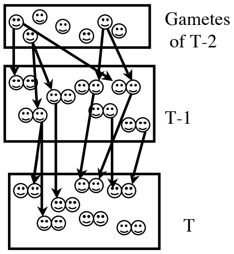
Genetic Drift vs. Inbreeding

- Drift autozygosity
  - Limited population size and random mating
    - $F_t \neq 0$
    - Some individuals will be autozygous by chance.



Genetic Drift vs. Inbreeding

- Drift autozygosity
  - Limited population size and random mating
    - $F_t \neq 0$
    - Some individuals will be autozygous by chance.



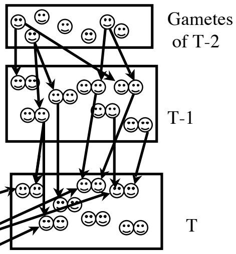
Allozygous/homozygous

Autozygous/homozygous

Allozygous/heterozygous

Drift vs. Inbreeding

- Drift autozygosity
  - Limited population size and random mating
    - $F_t \neq 0$
    - Some individuals will be autozygous by chance.



Allozygous/homozygous

Autozygous/homozygous

Allozygous/heterozygous

Genetic Drift vs. Inbreeding

- Drift autozygosity
  - Limited population size and random mating
    - $F_t \neq 0$
    - Some individuals will be autozygous by chance

Probability of being autozygous in this generation	1/2N F=1
Probability of being autozygous from the previous generation	1-1/2N $F = F_{t-1}$

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

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

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

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

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

Genetic Drift vs. Inbreeding

- F = inbreeding coefficient or the probability of autozygosity based on finite population size within individual subpopulations.
- The smaller the population size the greater the amount of genetic drift a population experiences
- The more genetic drift a population experiences the greater the inbreeding (F)

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

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

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

$$1 - F_{t-2} = \left(1 - \frac{1}{2N}\right)(1 - F_{t-3})$$

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

If we assume that $F_0 = 0$, then

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

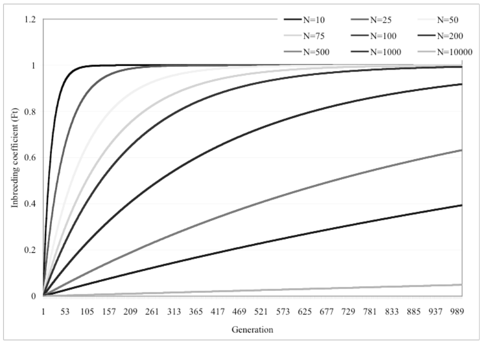
Inbreeding and HWE

- As a population becomes more inbreed we expect to see the population becoming more homozygous over time.
  - Ways to create homozygotes
    - Start with an A_1 egg and pair it with an A_1 sperm that is not IBD = p*p*(1-F)
      - Allozygous homozygote
      - No common ancestry
      - Multiple mutations or migration
    - Start with an A_1 egg and pair it with an A_1 sperm that is IBD = p(F)
      - Autozygous homozygote
      - Common ancestor
      - Breeding

$$A_1A_1 \Rightarrow p^2(1-F) + pF$$

$$A_1A_2 \Rightarrow 2pq(1-F)$$

$$A_2A_2 \Rightarrow q^2(1-F) + qF$$



Genetic Drift

- Assumptions
  - Diploid
  - Sexual reproduction
  - Nonoverlapping generations
  - Many independent subpopulations
  - Each subpopulation has a constant size of N
  - Random mating
  - No migration
  - No mutation
  - No selection

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

If this gives the rate of increase in autozygosity, heterozygotic decay is

$$\mathrm{H_t} = 1 - F_t$$

Analagously to what we did with F

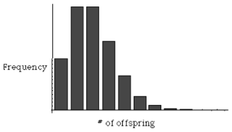
$$\mathrm{H_t} = \left(1 - \frac{1}{2N}\right) H_{t-1}$$

Using the recursion equation

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

Effective Population Size

- N_e - The size of an ideal population that loses genetic variation at the same rate as the focal population
  - Ideal Population
    - Non-overlapping generations
    - Equal sex ratio
    - Poisson variance in reproductive success
    - Constant population size
    - No isolation by distance
- Under these conditions populations will lose heterozygosity at a constant rate and will reach an equilibrium with mutation.



Effective Population Size

- N_e - The size of an ideal population that loses genetic variation at the same rate as the focal population
  - Ideal Population
    - Non-overlapping generations
    - Equal sex ratio
    - Poisson variance in reproductive success
    - Constant population size
    - No isolation by distance
- Under these conditions populations will lose heterozygosity at a constant rate and will reach an equilibrium with mutation.
  - Drift removes alleles at the same rate mutation is adding them.

Infinite Alleles Model

$$H = \frac{4N_e\mu}{4N_e\mu - 1}$$

$$N_e = \frac{H}{4\mu(1-H)}$$

Stepwise Mutation Model

$$H = 1 - \frac{1}{(8N_e\mu - 1)^{0.5}}$$

$$N_e = \left(\frac{1}{8\mu}\right)\left(\left(\frac{1}{1-H}\right)^2 - 1\right)$$

Effective Population Size

- Given the definition of N_e
  - The size of an ideal population that loses genetic variation at the same rate as the focal population
- You can also estimate N_e directly based on the genetic change between generations.
- Change in heterozygosity

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

$$\frac{H_t}{H_0} = \left(1 - \frac{1}{2N_e}\right)^t$$

$$\left(\frac{H_t}{H_0}\right)^{\frac{1}{t}} = 1 - \frac{1}{2N_e}$$

$$-1 + \left(\frac{H_t}{H_0}\right)^{\frac{1}{t}} = -\frac{1}{2N_e}$$

$$1 - \left(\frac{H_t}{H_0}\right)^{\frac{1}{t}} = \frac{1}{2N_e}$$

$$2\left(1 - \left(\frac{H_t}{H_0}\right)^{\frac{1}{t}}\right) = \frac{1}{N_e}$$

$$\frac{1}{2\left(1 - \left(\frac{H_t}{H_0}\right)^{\frac{1}{t}}\right)} = N_e$$

N_e

- It is also possible to estimate N_e based on the temporal change in allele frequencies between generations
  - EQ.1 - Estimate of temporal change is the variance in allele frequencies between 2 time periods
  - EQ.2 - N_e is inversely related to the amount of variance. The $1/2S_i$ terms account for variance due to sampling error (our sampling).
    - Allows selfing
  - EQ.3 - Estimate of N_e
    - without selfing
  - EQ.4 - Estimate of N_e
    - $N = N_e$
    - Without selfing
  - EQ.5 - Estimate of a if $N \neq N_e$
- Multiple loci are incorporated by averaging F_k across loci

Eq. 1 $$\mathrm{F_k} = \frac{1}{A-1}\sum \frac{(x_i - y_i)^2}{(x_i + y_i)/2}$$

Eq. 2 $$N_e = \frac{t}{2\left[F_k - \frac{1}{2S_0} - \frac{1}{2S_t} + \frac{1}{N}\right]}$$

Eq. 3 $$N_e = \left(\frac{1}{2(1 - e^a)}\right) - 0.5$$

Eq. 4 $$a = \frac{\ln\left(1 - F_k - \frac{1}{2S_0} - \frac{1}{2S_t}\right)}{t} \text{ if } \mathrm{N = N_e}$$

Eq. 5 $$a = \frac{\ln\left(1 - F_k - \frac{1}{2S_0} - \frac{1}{2S_t} + \frac{1}{N}\right)}{t} \text{ if } \mathrm{N \neq N_e}$$

A = # alleles

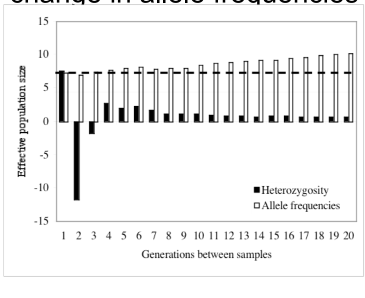
x_i = frequency of allele i in 1st period

y_i = frequency of allele i in 2nd period

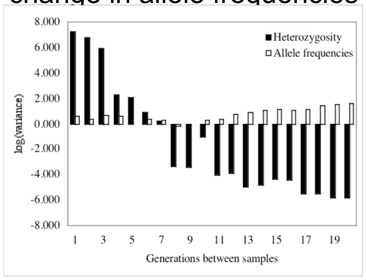
S_i = # individuals sampled in period i

N = # of ind in Pop.

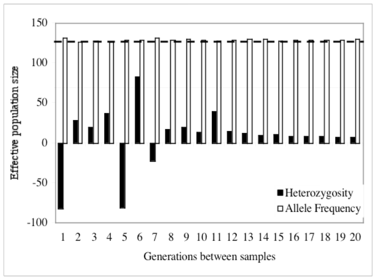
Change in H vs. change in allele frequencies



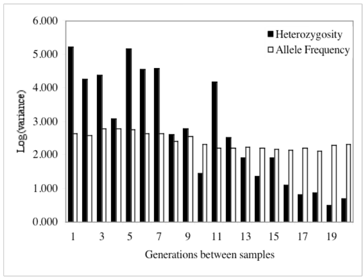
Change in H
vs.
change in allele frequencies



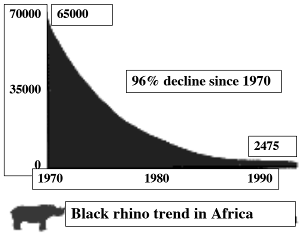
Change in H vs.
change in allele frequencies



Change in H vs.
change in allele frequencies



Black rhino
(*Diceros bicornus*)



Method	Formula	Eq. #	Citation
SR	$N_e = \frac{4N_m N_f}{N_m + N_f}$	1	Wright 1969
FPS	$\frac{1}{N_e} = \frac{1}{t}\left(\sum_{i=1}^{t} \frac{1}{N_i}\right)$	2	Wright 1969
VRS	$N_e = \frac{4N - 2}{\sigma_k^2 + 2}$	3	Wright 1969
SRFPS	$\frac{1}{N_e} = \frac{1}{t}\left(\sum_{i=1}^{t} \frac{1}{N_{SR,i}}\right)$	4	Chepko-Sade *et al*. 1987
SRVRS	$N_e = \frac{4N_{mVRS}N_{fVRS}}{N_{mVRS} + N_{fVRS}}$	5	Chepko-Sade *et al*. 1987
FPSVRS	$\frac{1}{N_e} = \frac{1}{t}\left(\sum_{i=1}^{t} \frac{1}{N_{VRS,i}}\right)$	6	Chepko-Sade *et al*. 1987
TRI	$\frac{1}{N_e} = \frac{1}{t}\left(\sum_{i=1}^{t} \frac{1}{N_{SR+VRS,i}}\right)$	7	Chepko-Sade *et al*. 1987

N_e Fluctuating Population Size

- Natural populations change in size over time
- Small populations lose genetic variation faster than large ones
- Once variation is lost it takes a long time to recover it
- Short periods of small population size have a greater effect than long periods of large populations size
  - Use the Harmonic mean rather than arithmetic mean

$$1 - F_2 = \left(1 - \frac{1}{2N_1}\right)\left(1 - \frac{1}{2N_0}\right)(1 - F_0)$$

And since

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

$$1 - F_2 = \left(1 - \frac{1}{2N}\right)^2 (1 - F_0)$$

But, N must be the effective size of the varying population sizes $\mathrm{N_e}$

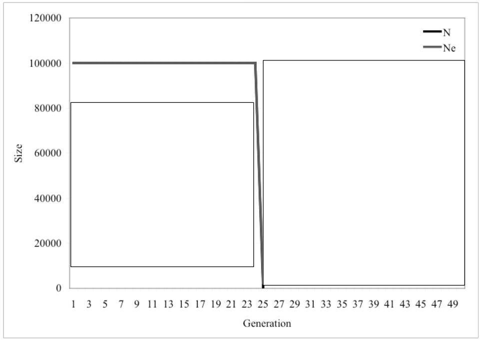
$$\left(1 - \frac{1}{2N}\right)^2 (1 - F_0) = \left(1 - \frac{1}{2N_1}\right)\left(1 - \frac{1}{2N_0}\right)(1 - F_0)$$

$$\left(1 - \frac{1}{2N}\right)^2 = \left(1 - \frac{1}{2N_1}\right)\left(1 - \frac{1}{2N_0}\right)$$

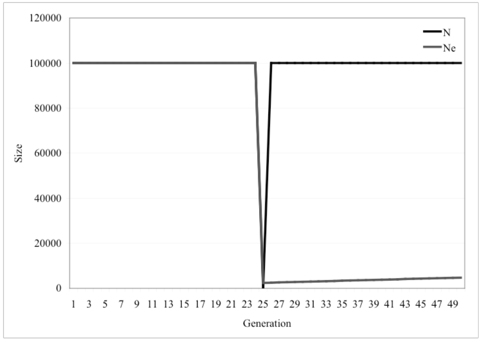
Which is approximated by

$$\frac{1}{\mathrm{N_e}} = \frac{1}{2}\left(\frac{1}{N_0} + \frac{1}{N_1}\right) \Rightarrow \frac{1}{\mathrm{N_e}} = \frac{1}{t}\left(\frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N_t}\right)$$

N_e Fluctuating Population Size



N_e Fluctuating Population Size



Black-footed ferret
(*Mustela nigripes*)

N_e Unequal Sex Ratio

- The sex ratio is unlikely to be 1:1 (but also unlikely to be far from 1:1, why?)
- In species with separate sexes the probability that two alleles comes from the same sex is 1/2*1/2
- The probability that two alleles come from the same female is $1/2N_F*1/2N_F$
  - Similarly for males $1/2N_M*1/2N_M$
  - $1/4N_F$ and $1/4N_M$
- The derivation is finished to the right.

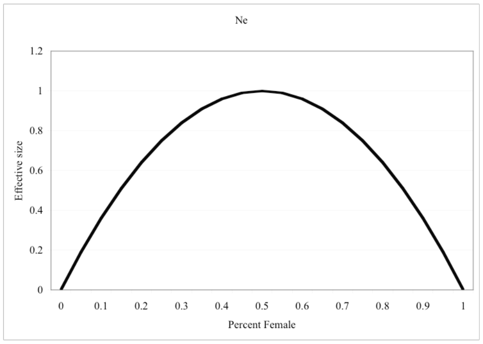
$$\frac{1}{\mathrm{N_e}} = \left(\frac{1}{2N_F} * \frac{1}{2N_F}\right) + \left(\frac{1}{2N_M} * \frac{1}{2N_M}\right)$$

$$\frac{1}{\mathrm{N_e}} = \frac{1}{4N_F} + \frac{1}{4N_M}$$

$$\frac{1}{\mathrm{N_e}} = \frac{N_M}{4N_F N_M} + \frac{N_F}{4N_F N_M}$$

$$\frac{1}{\mathrm{N_e}} = \frac{N_M + N_F}{4N_F N_M}$$

$$N_e = \frac{4N_F N_M}{N_M + N_F}$$



N_e Homogametic genes

- Sex chromosomes
  - X linked genes are transferred based on the sampling variance for the X chromosome in males and females

$$N_e = \frac{9N_M N_F}{4N_M + 2N_F}$$

Variances $\frac{p_m q_m}{N_m}$ $\frac{p_f q_f}{2N_f}$

Frequency of X in population $p = \frac{1}{3}p_m + \frac{2}{3}p_f$

$$Var(p) = \frac{1}{9}\left(\frac{p_m q_m}{N_m}\right) + \frac{4}{9}\left(\frac{p_f q_f}{2N_f}\right)$$

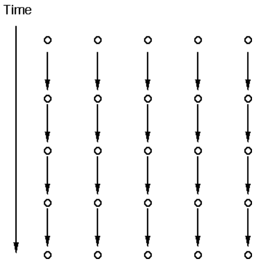
at equilibrium $p_m = p_f$,*thus*

$$Var(p) = pq\left(\frac{1}{9}\frac{1}{N_m} + \frac{4}{9}\frac{1}{2N_f}\right) = \frac{pq}{2\left[\frac{9N_m N_f}{4N_m + 2N_f}\right]}$$

$$Var(p) = \frac{pq}{2N_e}$$

Elephant seal
(*Mirounga angustirostris*)

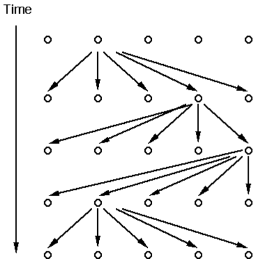
Variation in Rep. Success



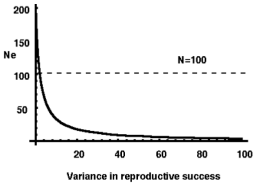
- No Variation is achieved when all alleles have exactly the same number of offspring alleles.
  - $F_t=F_{t-1}$
  - $F_t=(1/(2N_e))F_{t-1}+1/(2N_e)$
  - N_e = infinity

Variation in Rep. Success

- Maximum Variation is achieved when all offspring arise from one parent.
  - $F_t=1$
  - $F_t=(1/(2N_e))F_{t-1}+1/(2N_e)$
  - N_e = 1/2



Variation in Rep. Success



$$N_e = \frac{4N-2}{\sigma_k^2+2}$$

- The variance in reproductive success among individuals with a Poisson distribution is 2 in a stable population
  - $N_e = N - 0.5$

Variance N_e

- N_e can be thought of as being related to the amount of genetic change in a population, or the amount of variation in allele frequencies.
- Since the variance associated with an ideal population is given by the top equation we can just rearrange the equation to be able to solve for N_e based on the variance in allele frequencies over time

$$\mathrm{var}(p) = \frac{p(1-p)}{2N_e}$$

$$N_e^v = \frac{p(1-p)}{2\,\mathrm{var}(p)}$$

$$F_t = \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right)F_{t-1}$$

$$F_t = \frac{1}{2N_e} - \frac{F_{t-1}}{2N_e} + F_{t-1}$$

$$F_t = \frac{1}{2N_e}(1 - F_{t-1}) + F_{t-1}$$

$$F_t - F_{t-1} = \frac{1}{2N_e}(1 - F_{t-1})$$

$$N_e = \frac{1 - F_{t-1}}{2(F_t - F_{t-1})}$$

$$\text{if } \mathrm{F}_{t\text{-}1} = 0$$

$$N_e = \frac{1}{2F_t}$$

Inbreeding effective size

- Since the effective population size is related to the loss of genetic variation it is also related to the increase in the amount of inbreeding
- The assumption $F_{t-1}= 0$ is OK because we are concerned with how much F changes, not the magnitude of the change

Comparing N_e

$$N_e^f = \frac{1}{2F_t}$$

vs.

$$N_e^v = \frac{p(1-p)}{2\,\mathrm{var}(p)}$$

$$N_e^f \neq N_e^v$$

$$\text{if } N_1 \neq N_2 \neq N_3 \neq N_4 \neq N_4$$

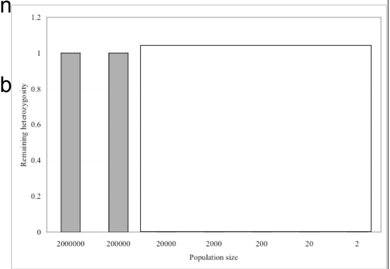
- Inbreeding effective size is naturally related to the number of individuals in the parental populations.
  - It tells you something about how the probability of identity by descent within a single population will change over time.
- Variance effective size is naturally related to the number of individuals in the offspring generation.
  - It tells you something about how much allele frequencies in isolated populations will diverge from one another.

Heterozygosity and Bottlenecks

- Each generation genetic drift reduces heterozygosity b

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1}$$

$$\frac{H_t}{H_{t-1}} = \left(1 - \frac{1}{2N}\right)$$

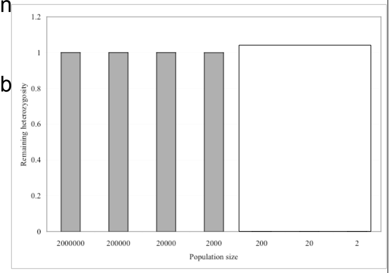


Heterozygosity and Bottlenecks

- Each generation genetic drift reduces heterozygosity b

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1}$$

$$\frac{H_t}{H_{t-1}} = \left(1 - \frac{1}{2N}\right)$$

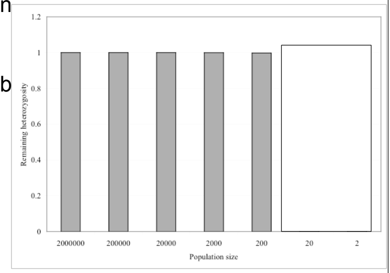


Heterozygosity and Bottlenecks

- Each generation genetic drift reduces heterozygosity b

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1}$$

$$\frac{H_t}{H_{t-1}} = \left(1 - \frac{1}{2N}\right)$$

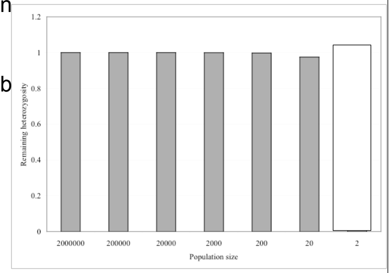


Heterozygosity and Bottlenecks

- Each generation genetic drift reduces heterozygosity b

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1}$$

$$\frac{H_t}{H_{t-1}} = \left(1 - \frac{1}{2N}\right)$$

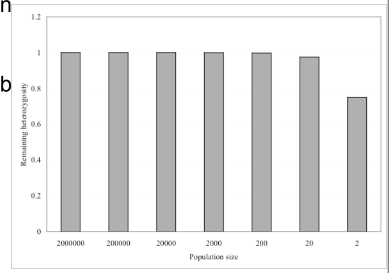


Heterozygosity and Bottlenecks

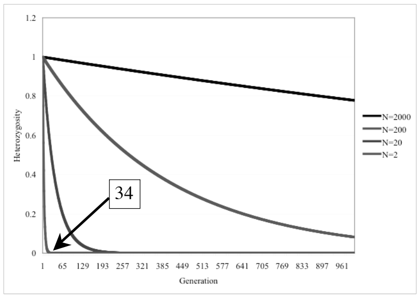
- Each generation genetic drift reduces heterozygosity b

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1}$$

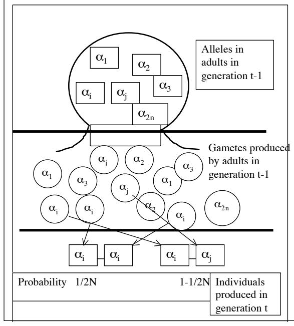
$$\frac{H_t}{H_{t-1}} = \left(1 - \frac{1}{2N}\right)$$



Heterozygosity and Bottlenecks

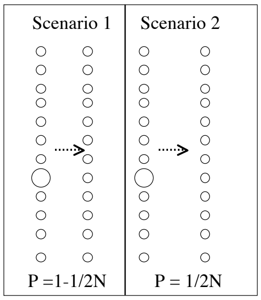


Behavior of a new mutant eventually in a population



- The frequency of a mutation is based on an equilibrium of μ and ν in an infinite population
- More realistic model doesn't assume ∞ N
  - Allele frequencies are based on mutation and Random Genetic Drift (RGD)

Probability of fixing a new mutant



- The probability an allele goes extinct any generation is 0.368
- The probability a new mutation gets fixed is 1/2N
  - Probability of fixation is random draw of the alleles so every one has an equally bad chance.
- Probability of losing an allele
  - Pr(losing A) = $(1-p)^{2N}$
  - This differs from the transition probability of going to 0 in that this is the probability of losing this allele at some point in the future

Probability of losing an allele each generation

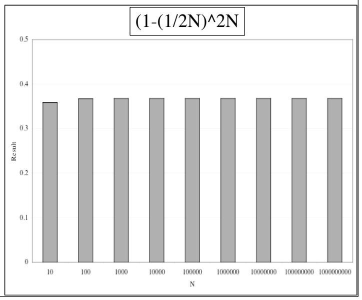
- RGD results in some alleles being under-represented in subsequent generations.
- The extreme of which is no representation
  - Extinction of the allelic lineage
- Probability of this is based on expansion of

$$\left[\frac{1}{2N}\alpha_1+\left(1-\frac{1}{2N}\right)\alpha\right]^{2N}$$

W f α 1 alleles and α is the number of non- α_1 alleles
- Which solves to

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

- by

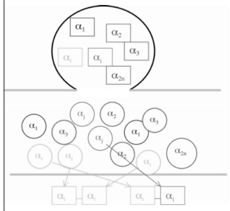


Infinite Alleles Model (IAM)

- Mutations
  - Unlikely that only 2 alleles exist.
  - How many alleles exist?
    - Lots
      - a protein with 300 AA
      - 900 nucleotides occupied by 1 of 4 nucs (A,T,C,g)
      - $4^{900} = 10^{542}$!!!
      - Typical microsatelite locus has 10 - 20 alleles
      - Typical allozyme locus has 2-3
      - Typical RAPD or AFLP?
  - Each mutation should create a novel allele thus there are an infinite number of possible alleles

- Mutation
  - Always creates a novel allele
    - Any two identical alleles must be Identical By Descent (IBD)
  - Calculating homozygosity
    - Two alleles that must be IBD join in an individual
      - Autozygous

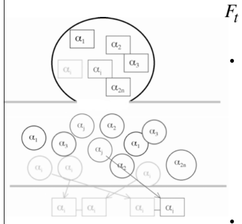
Infinite Alleles Model (IAM)



- Probability of autozygosity
  - F_t= prob. 2 alleles in generation t are IBD
  - Two way to be IBD
    - α_i α_i
      - Probability neither mutates
        - » $(1-\mu)^2$
      - Probability α_i α_l =1/(2N)
        - » Combine 1/(2N) $(1-\mu)^2$
        - » What is this process?
    - α_i α_j (in this case I =, ≠ j)
      - Probability neither mutates
        - » $(1-\mu)^2$
      - Alleles were IBD in t-1
        - » F_{t-1}
        - » F_{t-1} $(1-\mu)^2$
        - » Now go back through time: recursion equation
        - » Probability α_i α_j =1-1/(2N)
        - » Combine (1-1/(2N)) $(1-\mu)^2F_{t-1}$

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

Infinite Alleles Model (IAM)



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

- In this model RGD and mutation are working in opposite directions
  - Equilibrium
    - RGD - Increases autozygosity
    - Mutation - Decreases autozygosity
  - Do the math

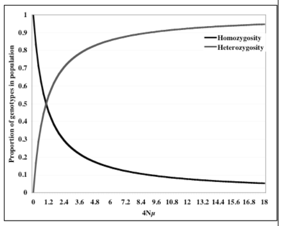
$$\hat{F} = \frac{1}{1+4N\mu}$$

- The expected amount of autozygosity/homozygosity at equilibrium with mutation.
- Heterozygosity = $\frac{4N\mu}{(4N\mu+1)}$

Infinite Alleles Model (IAM)

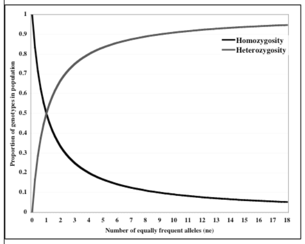
- Consequences
  - With mutations there is a narrow range over which you expect ≤ 50% homozygotes
- Notice F does not depend on the number of alleles!
- Two ways to calculate homozygosity

$$\hat{F} = \frac{1}{1+4N\mu}$$



$$\hat{F} = \frac{1}{1+4N\mu} = \sum_{i=1}^{n} p_i^{\,2} = p_1^2 + p_2^2 + \dots p_n^2$$

Infinite Alleles Models (IAM)



- Complication
  - Any number of allele distributions can produce the observed levels of homozygosity based on how many alleles are in the system.
  - Effective Number of Alleles (n_e)
    - The number of equally frequent alleles required for the observed amount of homozygosity
      - $p_1 = p_2 = \dots = p_n = 1/n$

$$\hat{F} = \sum_{i=1}^{n} p_i^{\,2} = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}$$

$$\frac{1}{n} = \hat{F} = \frac{1}{4N\mu+1} \Rightarrow n_e = 4N\mu+1$$

Neutral Theory

- The probability of fixing a neutral allele is $p_0 = 1/2N$

The equilibrium rate of fixation is μ.

Probability of fixing an allele = 1/2N
Number of mutations each generation = $2N\mu$
Thus $2N\mu/2N = \mu$

The average time between mutations = $1/\mu$

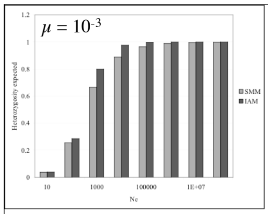
- The average time to fixation is $4N_e$
  - Neutral mutations take a long time to fix
  - Neural mutations are lost quickly

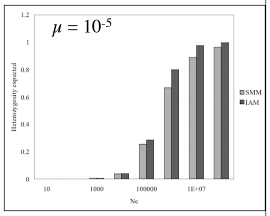
Under the Infinite Alleles Model
Equilibrium homozygosity = $1 / (4N_e\mu)$
Equilibrium heterozygosity =

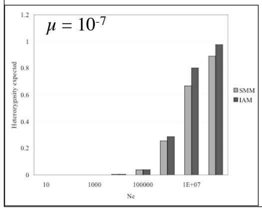
$$H = 1 - \frac{1}{4N_e\mu} = \frac{4N_e\mu}{1+4N_e\mu}$$

Under the Stepwise Mutation Model
Equilibrium heterozygosity

$$H = 1 - \left(\frac{1}{(8\mu N_e+1)^{0.5}}\right)$$







Why would different populations fix different alleles?