

# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

## DEFINITIONS

*Genetic drift*: Random fluctuation of allelic frequencies in a population of finite size.

*Continuous drift*: Genetic drift in permanently small populations.

*Bottleneck*: Drift due to a momentary restriction in population size.

*Founder events*: Drift that occurs when a limited by the number of founders colonize an isolated location.

## BINOMIAL SAMPLING

Reproduction can be viewed as a random sampling process. The progeny are a sample of the parents' gene pool.

$$(p + q)^a = 1, \text{ where } p + q = 1.$$

The variance of a binomial is

$$\sigma_p^2 = \frac{p(1-p)}{N}$$

Imagine a situation in which the frequency of the A allele is  $p = 0.5$  and only 2 parents will produce the next generation of progeny. There are  $2N$  genes = 4 sampled for the next generation, so

$$(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

What is the probability of getting no A alleles in the progeny?

$$p^4 = (0.5)^4 = 1/16$$

What is the probability of getting exactly 50% A alleles in the progeny?

$$6p^2q^2 = 6(0.5)^2(0.5)^2 = 0.375$$

More generally, we can solve such questions with the equation

$$E(n, N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$$

What is the probability of getting 50% A alleles from a sample of 5 parents?

$N = 10$ ,  $n = 5$ , and  $p = 0.5$ .

$$E(5, 10) = \frac{10!}{5!(5)!} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{3,628,800}{120 \cdot 120} \cdot \frac{1}{32} \cdot \frac{1}{32} = 252 \cdot \frac{1}{1024} = 0.246$$

That means that only about 25% of the time are you going to get 0.5 A alleles in a sample of ten gametes, and about 75% of the time you are going to get a biased sample (i.e., not 0.5).

# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

The variance around  $p$  in samples of 5 parents would be

$$\sigma_p^2 = \frac{p(1-q)}{2N} = \frac{0.5(1-0.5)}{10} = \frac{0.25}{10} = 0.025$$

In samples of 2 parents, it would be

$$\sigma_p^2 = \frac{0.5(1-0.5)}{4} = \frac{0.25}{4} = 0.0625$$

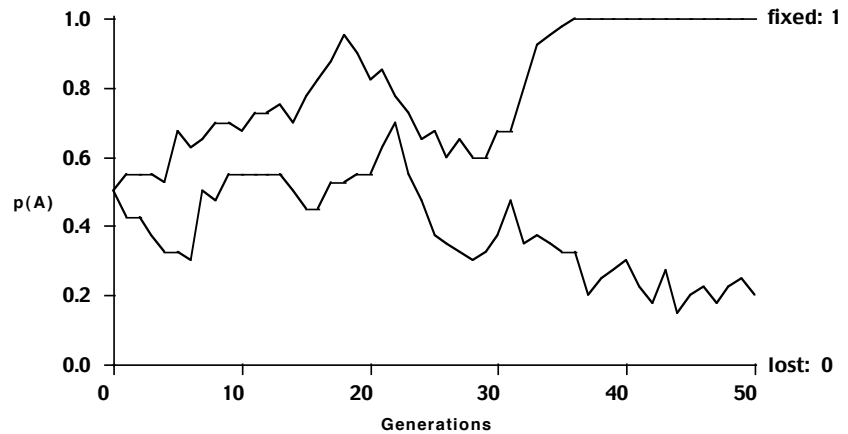
And in samples of 500 parents would be:

$$\sigma_p^2 = \frac{0.5(1-0.5)}{1000} = \frac{0.25}{1000} = 0.00025$$

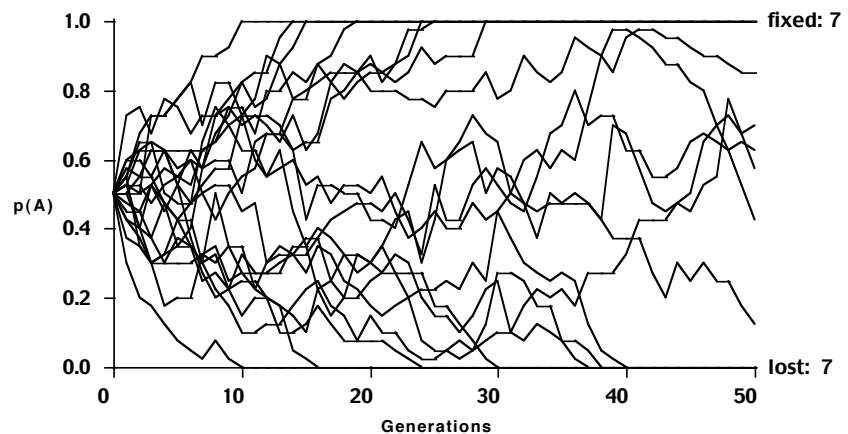
Random drift is a sampling problem in small populations or when there are a small number of parents contributing to the next generation

Continuous drift results in a “random walk” in the frequency of an allele. In the adjacent example the population consists of 10 males and 10 females in each generation. As you can see,  $p$  randomly changes in each generation and eventually the allele is either **fixed** or lost from the population.

Simulations were generated by PopGen 1.0: <http://cc.oulu.fi/~jaspi/popgen/popgen.htm>



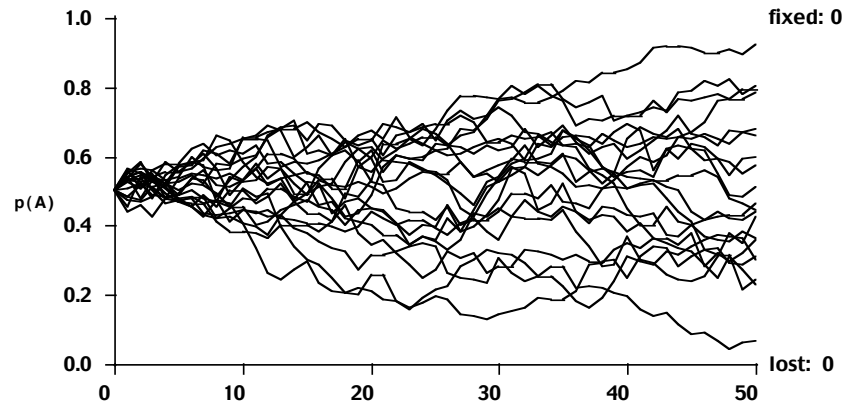
If the simulation were run many times (20 in the adjacent example with  $N_e = 20$ ), there would be equal numbers of cases in which  $p$  increases decreases. At the end of 25 generations, the mean  $p$  across these simulations would still equal  $p_0$ .



# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

However, if we run the simulation with  $N_e = 100$ , the rate of drift is significantly decreased. In addition, the variance in  $p$  in generation 50 is greatly constrained compared to the previous case. We get back to this issue of  $V_{pt}$  later.



## LOSS OF HETEROZYGOSITY

If we considered each simulation an independent gene locus, then it is obvious that the net effect of drift over many generations in a single population is to decrease the average heterozygosity across multiple gene loci.

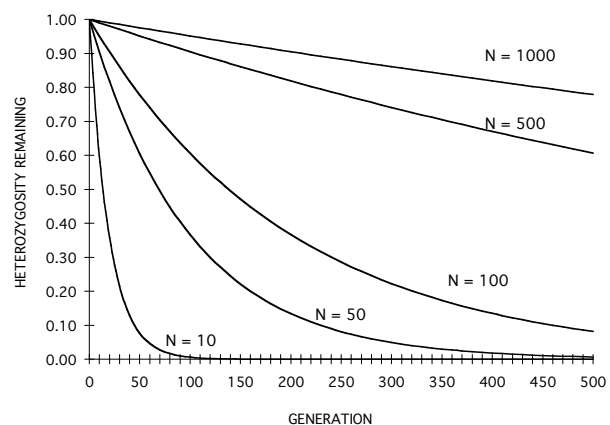
We can derive expected heterozygosity ( $H_t$ ) after  $t$  generations of drift as follows:  $H_0$  is heterozygosity in generation zero, before drift; there are  $N_e$  breeding adults in the population ( $N_e$  is the genetically effective size of the population); and each individual (diploid) contributes two gametes to the next generation.

$$H_t = H_0 \left(1 - \frac{1}{2N_e}\right)^t, \text{ which can be approximated by } H_t = H_0 e^{\frac{-t}{2N_e}}$$

The rate of loss of  $H_t$  is a function of  $N_e$ , the number of breeding adults in the population. We saw this same relationship under inbreeding.

You can solve this equation for the number of generations it takes to reduce heterozygosity by a certain amount

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



What is the half-life of heterozygosity ( $H_t/H_0 = 0.5$ )?

$$t = -2N_e \ln\left(\frac{H_t}{H_0}\right) = -2N_e \ln\left(\frac{1}{2}\right) = -2N_e \times -0.69 = 1.39N$$

It takes about  $1.39N$  generations to reduce heterozygosity by 1/2. If  $N_e = 10$ , it would take about 14 generations. If  $N_e = 100$ ,  $t \approx 139$  gen. If  $N_e = 1000$ ,  $t \approx 1390$  gen.

# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

How large must  $N_e$  be to maintain 90% of the heterozygosity for 10 generations?

$$\frac{H_t}{H_0} = e^{\frac{-t}{2N_e}} = 0.90, \text{ and } \ln(0.90) = \frac{-t}{2N_e}$$

Solving for  $N_e$  and substituting  $t = 10$ .

$$N_e = \frac{-t}{2 \ln(0.90)} = \frac{-10}{2 \ln(0.90)} = 47.5 \text{ generations.}$$

How large must  $N_e$  be to maintain 90% of the heterozygosity for 100 generations?

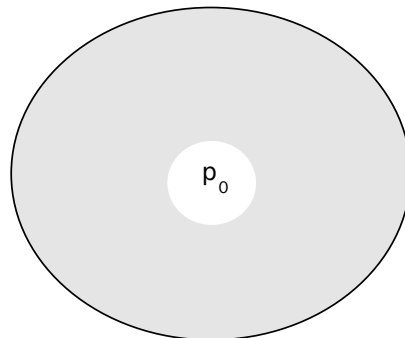
$$N_e = \frac{-100}{2 \ln(0.90)} = 475 \text{ generations.}$$

This is where some of the "magical" requirements you see for  $N_e$  in the conservation literature come from. As you can see, they are strictly dependant on the conditions you set for  $H_t/H_0$  and  $t$ . Those conditions are arbitrary. We might as well ask how big does  $N_e$  have to be to maintain 99% of the heterozygosity for 1000 generations. The answer is about 50,000. It is all based on a value judgment of how much variation you are willing to sacrifice.

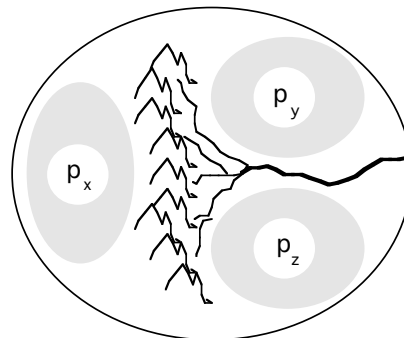
## FRAGMENTATION AND DIVERSIFICATION AMONG DEMES

Imagine starting with a large panmictic population and subdividing it into many small demes with no migration. Heterozygosity is lost as each deme drifts, and the variance in  $q$  among demes will grow.

before fragmentation



after fragmentation



$t_0$  —————  $n$  generations —————  $t_n$

Define:

$H_t$  as total genetic diversity (= heterozygosity),

$H_s$  as the diversity within demes, and

$V_q$  as the variance among demes.

# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

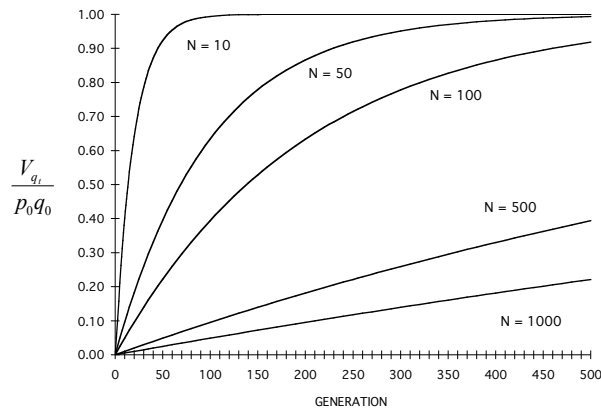
Total genetic diversity will be  $H_t = H_s + V_q$ .

$H_t$  is conserved, because  $H_s$  decreases while  $V_q$  increases. Gene diversity just shifts from within to among subpopulations.

$V_q$  grows according to the following relationship:  $V_{q_t} = p_0 q_0 \left[ 1 - \left( 1 - \frac{1}{2N_e} \right)^t \right]$

$$\text{and } \frac{V_{q_t}}{p_0 q_0} = 1 - \left( 1 - \frac{1}{2N_e} \right)^t,$$

which is the variance at generation  $t$  normalized by the theoretical maximum variance ( $p_0 q_0$ ) that would exist if all populations drifted to fixation for the alternative alleles? This equation is identical to the equation for  $f_i$  given during the inbreeding lecture. The term  $V_q/(p_0 q_0)$  as we will see in the subsequent lecture on population structure is treated like an inbreeding coefficient.



## GENETICALLY EFFECTIVE POPULATION SIZE

The genetically effective size ( $N_e$ ) of a population with the census size  $N$  is the size of an ideal panmictic population in which all parents have an equal probability of being the parents of any offspring. In natural age-structured populations this ideal is typically not the case. Several factors affect  $N_e$ .

### Uneven sex-ratios:

For species in which males have a harem (e.g., the northern elephant seal),  $N_e$  is strongly affected by sex-ratio bias.

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

For example, if each dominant male mates with 10 females.

$$N_e = \frac{4(10)(1)}{(10) + (1)} = \frac{40}{11} = 3.6 \text{ individuals}$$

even though the census size is  $N = 11$ . If harems contain 50 females, the effective size would only be  $N_e = 3.9$ . How would sneaky males affect this picture?

### Variance in family sizes

Just like variance in contributions of the two sexes affects  $N_e$ , excessive variance in family size among females will also diminish  $N_e$ . Some families will be represented better than others in the population will. Family size is Poisson distributed with a mean of

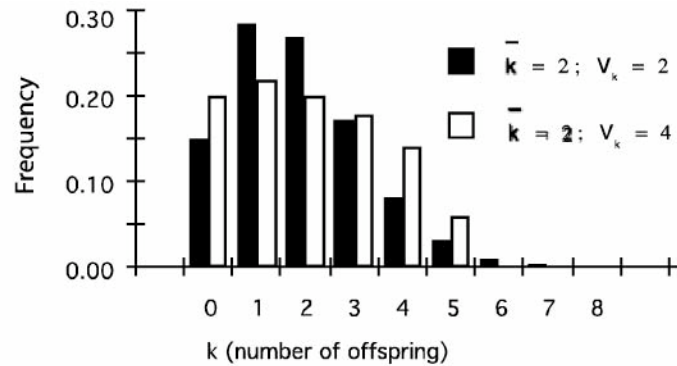
# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

$\bar{k}$  and a variance  $V_k$  that are equal. For a biparental species under stationary conditions, the effective size would be

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

In the lower panel, black bars represent a Poisson distribution with a variance to mean ratio ( $V_k / \bar{k}$ ) of 1.  $N_e \approx N$ . White bars represent a population with a  $V_k / \bar{k} = 2$ .  $N_e \approx 0.66N$ . If we equalize family sizes in a zoo population (i.e.,  $V_k = 0$ ) can we make  $N_e > N$ ? Yes, in this case  $N_e = 2N - 2$ . The ratio of  $N_e/N$  is approximately 2.



Declining populations in suboptimal habitats would have a mean family size smaller than 2 (the replacement rate) and a variance to mean ratio ( $V_k / \bar{k}$ ) that increases with the reductions in  $\bar{k}$ , further reducing  $N_e$ .

## Variance in colony size over time

If local populations vary in size over time,  $N_e$  will equal the harmonic mean of the  $N_i$ .

$$N_e = \frac{t}{\sum_{i=1}^t \frac{1}{N_i}}$$

Take a population of insects that ordinarily maintains about 1000 individuals. Lets follow two scenarios for ten generations:

Case	$N_i$ in generation						$\bar{N}$	$\tilde{N}$
	1	2	3	4	...	10		
A	1000	1000	1000	1000	...	1,000,000	100,900	1,250
B	1000	1000	1000	1000	...	10	901	48

In case A, the population explodes to one million insects in generation 10, and in case B is crashes to 10 individuals in generation 10. For comparison, the arithmetic means is given. The effective sizes are based on the harmonic mean of the  $N_i$ . Clearly, a bottleneck in population size (case B) has a drastic effect on  $N_e$ .

# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

## Age Structure

For most long-lived organisms like birds, mammals, and trees, the population is age-structured. Post-reproductive individuals do not contribute to  $N_e$ , but pre-reproductive individuals have a large *potential* contribution depending on the probability of surviving to reproductive age. Estimating  $N_e$  from an age-structured population with overlapping generations is complicated. A formal treatment is provided by Crow and Kimura (1970)

## Inbreeding

Inbreeding also reduces effective size. Crow and Kimura (1970) provide a more formal distinction between **variance effective size** (which is what we have been talking about up to now in this lecture) and **inbreeding effective size**. If  $f$  is the inbreeding coefficient,

$$N_e = \frac{N}{1 + f}$$

As you can see, if  $f = 0$ ,  $N_e = N$ . However, in a selfing population,  $f$  will approach one and  $N_e$  will equal  $N/2$ .

## Isolation by distance (neighborhood size)

In continuously distributed populations, individuals might be isolated from one another by distance and their dispersal abilities. Most individuals will mate with their neighbors. The idea of a genetic neighborhood was introduced by Wright (1943). Neighborhood size is defined as the number of individuals circumscribed by a circle with a radius that is twice the standard deviation ( $\sigma$ ) of the per generation migration distance. If you know the density ( $N_d$ ) of individuals in the circle defined by  $4\pi\sigma$ , then

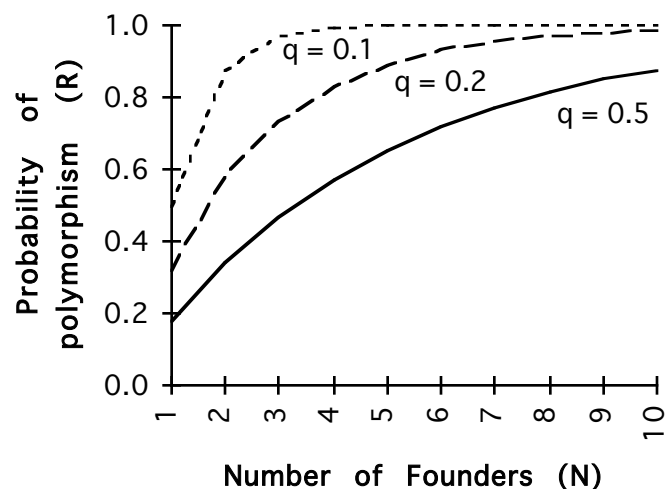
$N_e = 4\pi\sigma^2 N_d = 12.566\sigma^2 N_d$ . Wright (1943). showed that a circle of this size would include 86.5% of the parents of offspring at the center of this circle. We will deal with this concept again when we discuss migration and gene flow.

## FOUNDER EVENTS AND BOTTLENECKS

The importance of a founder event is a function of  $N$ , the founder size. The probability of retaining polymorphism can be expressed as:

$$R = 1 - p^{2N} - q^{2N}$$

Note that as allelic frequencies are more even, the probability of retaining polymorphism is higher. Clearly, rare alleles are most likely to be lost during founder events and population bottlenecks.



If a population grows quickly (high intrinsic rate of increase,  $r$ ) after a bottleneck or founder event, there will be little subsequent loss of heterozygosity. Nei et al. (1975)

# SMALL POPULATIONS AND GENETIC DRIFT

R. Vrijenhoek

showed that a founder event will eliminate alleles of low frequency, but the average heterozygosity is mostly influenced by  $r$ .

For example, from the relationship

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

we can see that even a founder event of  $N_e = 2$ , will only reduce heterozygosity to 75% of its former value.

$$\frac{H_t}{H_0} = \left(1 - \frac{1}{2 \cdot 2}\right)^1 = 1 - \frac{1}{4} = 0.75.$$

However, such an event could not leave more than 4 alleles at a locus. Sometimes founder events and bottlenecks result in rare alleles becoming more frequent. Imagine a locus with one common allele and many rare alleles. If  $N = 2$  and if one of the individuals ( $A_1 A_1 \times A_1 A_2$ ) just happened to be heterozygous for rare allele  $A_2$  in which  $q_0 \leq 0.01$ , then in one step  $q_0 \rightarrow q_1 = 0.25$ . So, even though many rare alleles are lost, some can be elevated by chance to higher frequencies. An example exists in the Amish people of central Pennsylvania. They grew from a founder population of about 200 individuals who migrated from Rhineland Germans between 1720 and 1770. Other immigrated subsequently, but as a group, they have remained isolated from the American populations around them. They have elevated frequencies of a number of generally rare genetic traits including Ellis-van Creveld syndrome which manifests as dwarfism, congenital heart defects, and polydactyly.

**Example of a bottleneck.** Pingelap atoll is Micronesian island that was struck by severe typhoon in 1775. Left about 30 survivors. Today has > 1600 people. A recessive disease, achromatopsia, a form of autosomal color blindness, occurs in about 5% of the natives ( $q \approx 0.23$ ). It is extremely rare on other Micronesian islands. Local explanation - the god Isoahpahu (god of night vision) fell in love with several local women. Pedigree analysis revealed that the chief in 1775, Mwahuele, was a carrier.

If he was the only carrier among the 30 natives, then  $q$  would have been  $1/60 = 0.014$  after the typhoon. Thus, the bottleneck alone should not have caused the increase to  $q = 0.23$ . However, as chief he had several wives and many children (apparently 14).

## REFERENCES

- Crow J, Kimura M (1970) *An Introduction to Population Genetics Theory* Harper and Row, New York.
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.
- Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.