Effective Population Size

As seen, finite population size has many important evolutionary consequences, such as increasing the variance of allele frequencies through time, increasing the amount of identity by descent, and causing the loss or fixation of alleles. As also shown, the rate at which these effects occur is inversely proportional to population size. However, in an idealized case, a more accurate statement is possible concerning the relation of these genetic parameters to population size. The idealized case refers to a diploid population of N hermaphrodictic, self-compatible, random mating adults with discrete generations and no age structure. All individuals contribute the same number of gametes on the average to the next generation, and the sampling variation in that number is described by a Poisson probability distribution (this is often an implicit rather than explicit assumption). Under these idealized conditions, it is possible to define very accurately how finite population size influences the rate of evolution by genetic drift for certain genetic parameters of interest.

Inbreeding Effective Size

Let the genetic parameter be $F$, the average probability of identity by descent in uniting gametes in the population. Suppose we start out with N individuals (satisfying all the idealized assumptions mentioned above) who are initially totally unrelated (i.e., $F(0)=0$). Suppose these N individuals produce N offspring for the next generation (generation 1). Since they are hermaphrodictic, self-compatible, and randomly mating, the chance of drawing 2 alleles that are ibd (identical by descent) is

$$\text{Prob.} = \frac{1}{2N} = F(1)$$

This result is obtained because a total of 2N gametes are drawn at random. For ibd, the same gamete type must be drawn twice. It makes no difference what the first gamete is, only that the second gamete be of the same type. Since the initial generation was not inbred, the only way to be ibd this generation is to have both gametes come from the same parental individual. Given one gamete has already been sampled, the probability of obtaining a second gamete from the same parent with sampling with replacement (the Poisson assumption, implicitly) is $1/N$. However, due to Mendelian segregation, only 1/2 of the sampled gametes from the same parent will be ibd, so the total probability of the second gamete matching the first is simply $1/(2N)$.

Now consider the second generation. The probability that two gametes will be ibd by random sampling is once again simply $1/(2N)$. Hence, the probability that they are not ibd due to random sampling in the current generation is $1-1/(2N)$. However, even if gametes are not ibd due to random sampling of the current
generation, they can still be ibd due to inbreeding in previous generations. Hence, the total prob. of ibd at generation 2 is:

\[ F(2) = \frac{1}{2N} + \left[ 1 - \frac{1}{2N} \right] F(1) \]

In words, the above formula means that Prob (ibd at gen. 2) = Prob(ibd due to sampling at gen. 2) + (prob not ibd due to sampling at gen.2)Prob(ibd due to inbreeding in the previous generation).

In general,

\[ F(t) = \frac{1}{2N} + \left[ 1 - \frac{1}{2N} \right] F(t-1) \]

Hence,

\[ H(t) = "heterozygosity at generation t" = 1-F(t) \]
\[ = 1-\frac{1}{2N} - \left[ 1 - \frac{1}{2N} \right] F(t-1) = \left[ 1 - \frac{1}{2N} \right] [1-F(t-1)] \]

(Note, this heterozygosity is not necessarily the observed heterozygosity nor the expected heterozygosity under random mating, because it regards anything as heterozygous that is not identical by descent; i.e., homozygotes that are identical by state are "heterozygotes" in this case. Hence, this heterozygosity is just the inverse of the pedigree F and is simply the probability that 2 alleles in the same individual are not identical by descent. “Heterozygosity”, like “inbreeding”, is one word with several meanings in population genetics, so be careful.)

Since 1-F(t-1) is simply H(t-1), the above equation reduces to

\[ H(t) = [1-1/(2N)]H(t-1). \]

The solution to this difference equation is simply

\[ H(t) = [1-1/(2N)]^t H(0) \]

or, if H(0)=1 (as assumed above), \[ F(t) = 1-[1-1/(2N)]^t. \]

In these equations, the rate at which F increases as a function of t is determined completely by N. However, what happens when the population deviates from all the idealized assumptions made above? No matter what, after t generations there will be some realized level of inbreeding, say F(t). Then, the inbreeding effective size (\(N_{ef}\)) is simply defined as that value which makes the following equation true:

\[ F(t) = 1-[1-1/(2N_{ef})]^{t} \]

That is, \(N_{ef} = 1/(2\{1-F(t)\}^{1/t})\)
The inbreeding effective size is simply the fudge factor needed to make the above equation true, that is, the number that allows the rate of accumulation of inbreeding due to drift to be the same as in an idealized population of size $N_{ef}$. Note that $N_{ef}$ is determined exclusively from $F(t)$ and $t$; it has no direct dependency upon the actual population size. Population geneticists have derived several equations that relate inbreeding effective size to actual population size under a variety of assumptions (see textbook). However, the primary definition of inbreeding effective size is that given above, and all other equations for inbreeding effective size are based on certain restrictive assumptions that may or may not be appropriate for a given real population.

If one has pedigree data on a managed population so that $F(t)$ can be calculated, then it is possible to determine the inbreeding effective size without the use of secondary equations. For example, the average probability of identity by descent in the Speke’s gazelle breeding herd in 1979, regarding the four founders as unrelated (and hence the reference generation), was 0.1283. The average number of generations of these animals from the founders was 1.7237 generations. Then, the inbreeding effective size of the 1979 herd is 6.5305, a number which is considerably less than the census size of 19 breeding animals. This low size is attributable to the founder effect: although there were 19 breeding animals available in 1979, their genes were derived from only four founders and therefore they accumulated inbreeding at a very fast rate, as indicated by the low inbreeding effective size. In 1979, Templeton and Read (1983) instituted a new management program that included avoidance of breeding near biological relatives. The first generation bred from the 19 animals available in 1979 consisted of 15 offspring with an average probability of identity by descent of 0.1490. Using this value for $F(t)$ and augmenting $t$ by 1, the above equation yields an inbreeding effective size of 8.6932. Note that the inbreeding effective size is still smaller than the census size, once again showing the persistent effects of the initial founding event. However, note that the inbreeding effective size increased (6.5305 to 8.6932) even though the census size decreased (19 to 15) when our inference is confined to the animals bred under the new program. This increase in inbreeding effective size reflects the impact of the avoidance of inbreeding (in a system of mating sense). In order to focus more directly upon the genetic impact of the Templeton and Read (1983) breeding program, one can define the 1979 herd as the reference population rather than the four founders. Note that the difference in average probability of identity by descent of the first generation born into the Templeton and Read (1983) program from that of their parents is 0.0207. Since the parental herd is now the reference, $t=1$, so the above equation yields the inbreeding effective size of the first generation of Speke’s gazelle born under this breeding program to be 21.0556. Note that the inbreeding effective size is greater than the census size of 15. This reveals yet another widespread fallacy; that effective population sizes have to be smaller than census sizes.

Note also that we now have two very different inbreeding effective sizes (8.6932 and 21.0556) for the same animals. The first size tells us how rapidly inbreeding has occurred both before and during the Templeton and Read (1983)
program, the later size tells us about the effects on average probability of identity by
descent after the inception of the breeding program. Hence, both numbers give
valuable information. However, it is commonplace in the population genetic
literature to refer to “the effective population size” as if there were only one
effective size for a population. As the Speke’s gazelle example clearly shows, there
can be many different effective sizes for the same population as a function of
different (and meaningful) reference populations. As we will now see, there are
also different types of effective size as well as different reference populations.

Variance Effective Size

Genetic drift also causes random deviations about the allele frequency of the
previous generation. These random deviations are commonly measured as the
variance in allele frequency induced by finite sampling. Suppose 2N gametes are
sampled out of a gene pool for the idealized population mentioned above. These
assumptions mean that the sample follows the binomial distribution. Let p be the
original frequency of A, and let x be the number (not frequency) of A’s in the
sample. Then with an idealized population of size N, the binomial probability of
x=X is

\[ \text{Prob}(x=X) = \binom{2N}{X} p^X q^{2N-X} \]

where q=1-p. The mean value of x in the binomial is 2Np and the variance of x is
2Npq. The allele frequency of A in the next generation is x/(2N), so the mean allele
frequency is Mean(x)/(2N)=p and the variance of allele frequency is
Var(x)/(2N)^2=pq/(2N). In non-ideal cases, define the variance effective size, \( N_{ev} \), to
be the value that makes the following equation true:

\[ \text{Var(allele frequency)} = \frac{pq}{(2N_{ev})}. \]

When dealing with the accumulation of variance over t generations, the
following equation is used:

\[ \text{Var(allele freq. after t generations)} = pq\{1-\left[1-1/(2N_{ev})\right]^t\} \]

where p is the allele frequency at generation 0.

This effective size measures how rapidly allele frequencies are likely to change
and how rapidly isolated subpopulations diverge from one another under genetic
drift.
Once again -- there is no such thing as "the effective size of a population."
Effective size is defined with respect to a genetic parameter of interest, and as the
parameter of interest changes, the effective size changes. Hence, a population can be
characterized by several different "effective sizes" simultaneously. For example,
consider a founder or bottleneck effect. Suppose we start with 1000 idealized
individuals with no inbreeding (F(0)=0). A bottleneck or founder event occurs such
that the next generation consists of only 4 individuals; produced by randomly
sampling the gametes of the previous generation. What are the inbreeding and
variance effective sizes of the founder generation?

Inbreeding effective size. Since gametes are randomly sampled from all 1000
parents to produce the 4 individuals of the founder generation, the probability that
two gametes that united to form a founder are derived from the same parent is
simply 1/1000. Because there is no inbreeding in the parents, the prob. that two such
gametes are ibd is 1/2. Hence, F(founder)=1/2000 which means that the inbreeding
effective size of the founder population is 1,000.

Variance effective size. Since only 8 gametes are sampled, the variance of p
in the founder generation is pq/8, which implies a variance effective size of the
founder population of 4. Hence, the same founder population has two very
different "effective sizes" for the genetic parameters of identity by descent versus
variance of allele frequencies.

In general, inbreeding effective size are primarily sensitive to the number of
parents and their reproductive characteristics (or other generations even more
remote in the past, depending upon the population of reference), whereas variance
effective size is primarily sensitive to the number of offspring and their attributes.
In non-ideal populations (that is, real populations), these numbers are generally
different. Hence, the phrase "the effective size of the population" is meaningless
unless the genetic parameter of interest is also specified.

Eigenvalue Effective Size

Another important genetic parameter is the rate at which alleles become fixed
or lost; that is, the rate at which allelic variability is lost at a locus. In an idealized
population, this occurs at a rate of 1/(2N) per generation. Hence, the proportion of
loci that are "polymorphic" (i.e., no fixation for one ibd allele class) should decrease
at a rate of [1-1/(2N)] per generation. In real populations, this rate of loss is
expressed as [1-1/(2N_e^λ)] per generation, or after t generations,  [1-1/(2N_e^λ)]^t
where N_e^λ is the eigenvalue effective size.

For example, in the 1982 Spekes' Gazelle population, it was determined that
86.53% of the founding ibd alleles were not fixed in the herd (this number can also
be calculated from pedigree data with the use of computer). Recall t=2.7237 in 1982.
Hence, we find the N_e^λ that satisfies
which turns out to be 9.6614. Recall that the inbreeding effective size for the same animals was (both with respect to the original 4 founders as the reference population) 8.6932. Hence, the herd is accumulating pedigree inbreeding at a faster rate than it is losing genetic diversity.

Another widespread misconception is that inbreeding causes a loss of genetic diversity. Inbreeding, in the system of mating sense, has no impact per se on gamete frequencies, and hence does not promote or retard the loss of genetic variation by itself. In the pedigree sense, it is important to keep a clear distinction between causal relationships (e.g., genetic drift causes a reduction in genetic variability, as measured by $N_e\lambda$) versus correlational relationships (e.g., genetic drift reduces genetic variability, as measured by $N_{ef}$, and increases inbreeding in the sense of average probability of identity by descent, as measured by $N_{ef}$, leading to a negative correlation between genetic variation and inbreeding). However, correlational relationships can be violated in particular instances. For example, population subdivision (to be discussed later in course) increases $N_{ef}$ but decreases $N_e\lambda$. Hence, in some realistic biological situations, factors that increase pedigree inbreeding result in greater retention of polymorphism than, say, random mating or avoidance of inbreeding.

EX) Lichen grasshoppers (Trimeratropis saxatilis) on MO glades

Can get accurate population numbers (census #’s) --- See 50-400 animals/glade pop. Yet, when there is much genetic variation. E.g., for mtDNA, the average heterozygosity/nucleotide is 0.014 within a glade population. Using an equation that will be developed later, this implies that $N_{ef} = 0.014/\mu$, where $\mu$ is the mutation rate per nucleotide. $\mu$ has been estimated to be $10^{-8}$ for insects, implying $N_{ef} = 1.4$ million. Even if the estimate of $\mu$ is off by a couple of orders of magnitude, it is obvious that the inbreeding effective is much, much larger than the census size.

On the other hand, there is much genetic differentiation between even nearby glade populations, such as on Proffit Mountain:

<table>
<thead>
<tr>
<th>4</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>

|-------|--------|

Mark/recapture studies show that grasshoppers do not disperse between glades and this is consistent with marked genetic differentiation. The PM glade populations had the following mtDNA variants and frequencies:
The above frequencies yield a \( \text{Var(allele frequency)}/pq \) of 0.02. If there is total isolation, we expect from the definition of variance effective size that:

\[
0.02 = \frac{\text{Var(allele freq.)}}{pq} = 1 - \frac{\bar{a}}{\bar{d}} - \frac{1}{N_{ev}} \frac{d}{\bar{d}}
\]

where \( t \) = number of generations of isolation. (\( N_{ev} \) is used rather than \( 2N_{ev} \) because mtDNA is effectively haploid). This implies (recall Taylor’s theorem)

\[
N_{ev} \cdot -t/\ln(0.02)
\]

We think \( t \) is about 100 in this case. There is strictly one generation per year. Before settlers of European ancestry settled the region, forest fires were a regular and important occurrence. These fires kept the glades clear (and therefore larger) and cleared out the undergrowth in the forests between the glades. It turns out, these grasshoppers don’t disperse through the "undergrowth" so fire prevention not only decreased the glade habitat and thus population sizes, but also decreased between-population dispersal (gene flow). With \( t = 100 \), you get \( N_{ev} = 26 \). Since this is for females only (mtDNA is maternally inherited), this shows that the variance effective size is only a bit smaller than the current census sizes -- a result not unexpected. The above \( N_{ef} \)'s and \( N_{ev} \)'s are also the effective sizes for the total population (all glades pooled together) because of the complete genetic isolation.

Now consider the eigenvalue effective size for a glade population. If we regard the ancestral gene pool as being estimated by the pooled gene pools of the current populations (recall, drift alone doesn’t change the average allele freq.), then the ancestral gene pool before fire suppression had 9 haplotypes. Today, there is an average of 4.6 alleles per population. So about half of the alleles have been lost. Hence:

\[
0.5 = 1 - \frac{\bar{a}}{\bar{d}} - \frac{1}{N_{ev}} \frac{d}{\bar{d}}
\]
so $N_{e1} = -100/\ln(0.5) = 144$ females. This is about the current census size of the populations. However, because there are so many glades, the probability of all glade populations losing an ancestral allele is close to 0, so the eigenvalue effective size of the total Ozark population is effectively infinite.

Hence, the different $N_e$'s differ by orders of magnitude. $N_{ef}$ is extremely large because the number of ancestors for the grasshoppers currently living on a glade is much, much larger than the glade population (the ancestors could come from many different glades prior to the suppression of forest fires). The $N_{ev}$ and $N_{e\lambda}$ for a glade are much more sensitive to present conditions and reflects the small current population sizes. The current apparently complete genetic isolation insures that the total population is accumulating F and var(allele freq.) at the same rate as individual glade populations, but this current isolation effectively preserves all genetic variation in the global population, making $N_{e\lambda}$ extremely large for the total population.

------------------------------------------------------------------------

**50/500 Rule:**
Conservation biologists want to preserve genetic variation. How big of a population does one need to

1. get enough variation for good evolutionary health
2. preserve variation over the long term

The 50/500 rule states one would need to begin with at least 50 individuals to start with enough variation. Then, the breeding program would need to increase the population size to at least 500 to keep the levels of genetic variation over time. This rule came from a sound, evolutionary perspective first proposed by Franklin (1980) but has been distorted and misconstrued. Franklin specified that the 50 individuals referred to the *inbreeding effective size* and the 500 to the *variance effective size*, NOT to actual census sizes.

Conservation workers have mistaken these numbers to be actual numbers of individuals rather than the appropriate effective sizes. This has led to some sometimes serious misapplications of population genetic theory. For example, some endangered species have been allowed to go extinct because there were fewer than 50 individuals remaining. Workers thought it would be a waste of time trying to save the species. However, if species had recently been of a greater population size or more genetically interconnected than at present, it could have very well had an inbreeding effective size much greater than its actual census size!

In the grasshopper example, $N_{ef}$ is at least tens of thousands if not millions even though the census population sizes are on the order of 100 animals. Therefore, five grasshoppers would have satisfied the 50 rule just fine.