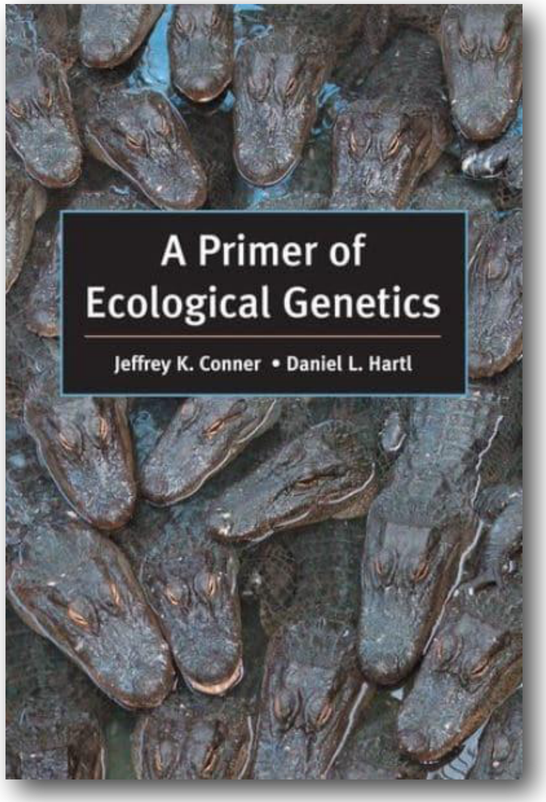
Evolutionary Genetics

LV 25600-01 | Lecture with exercises | 4KP

Effective Population Size

Effective Population Size



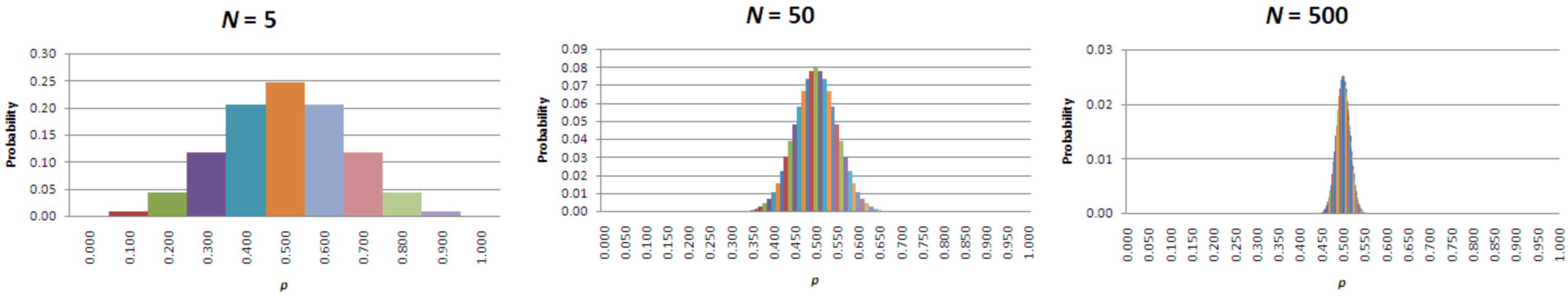
A Primer of Ecological Genetics

Jeffrey K. Conner • Daniel L. Hartl

Chapter 3: N_e (Pages 62-65)

PopGen ▷ Effective Population Size

Probabilities of allele frequencies in the next generation, in a population of 5, 50 and 500 organisms.

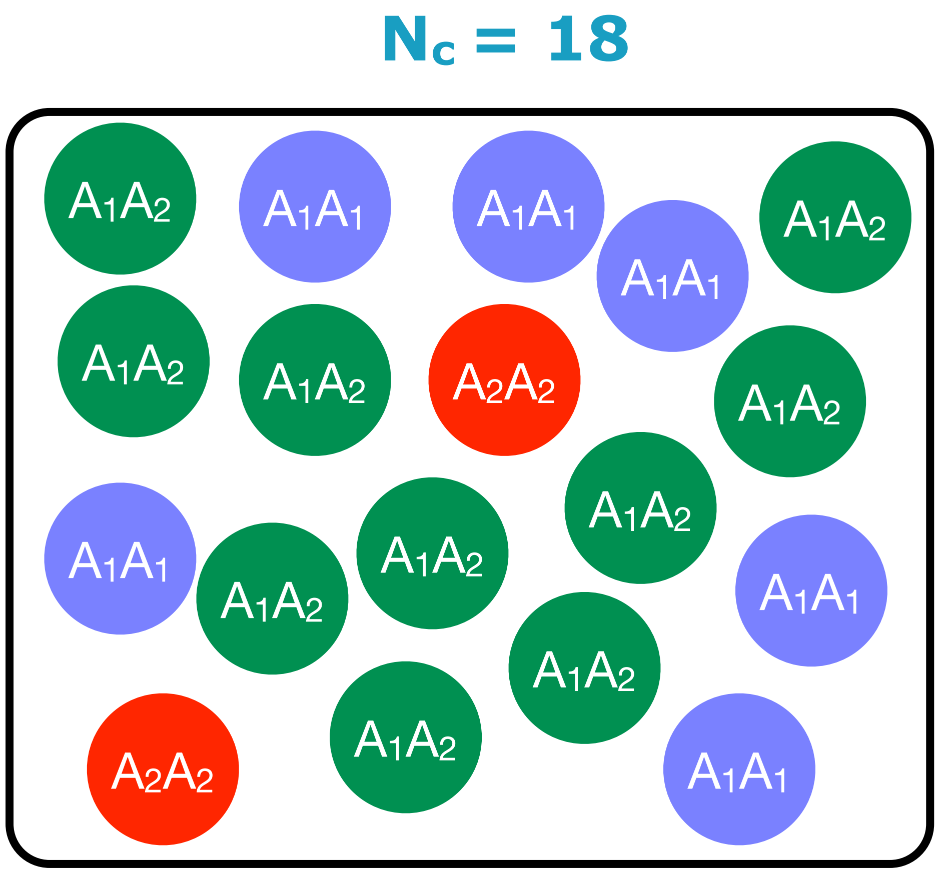


While allele frequencies are almost certain to change in each generation, **the amount of change due to sampling error decreases as the population size increases**. Perhaps most importantly, the **direction of change is unpredictable;** allele frequencies will randomly increase and decrease over time. Furthermore, when change does occur, sampling to produce the next generation will focus on the new value of p. Thus, given enough time, in the absence of factors that maintain both alleles (e.g. balancing selection), p will drift to either 0.0 or 1.0; in other words, one allele will drift to fixation and the other to extinction. The time it takes for this to happen depends on the initial frequencies of the alleles and, of course, on the **size of the population**.

Scitable by nature EDUCATION

PopGen ▷ Effective Population Size

Population size has a major influence on the dynamics of a population. For example, genetic drift reduces allelic diversity much faster in small populations than in large ones. Population size also affects the likelihood of extinction by randomly changing population size over time due to random variation in individual survival and reproductive success (demographic stochasticity). However, the rate of genetic drift is not really proportional to **census population size** (**Nc**). Rather, it's proportional to something more abstract — specifically, the **effective population size** (**Ne**).



An "ideal" population $N_c = N_e$ has the following characteristics

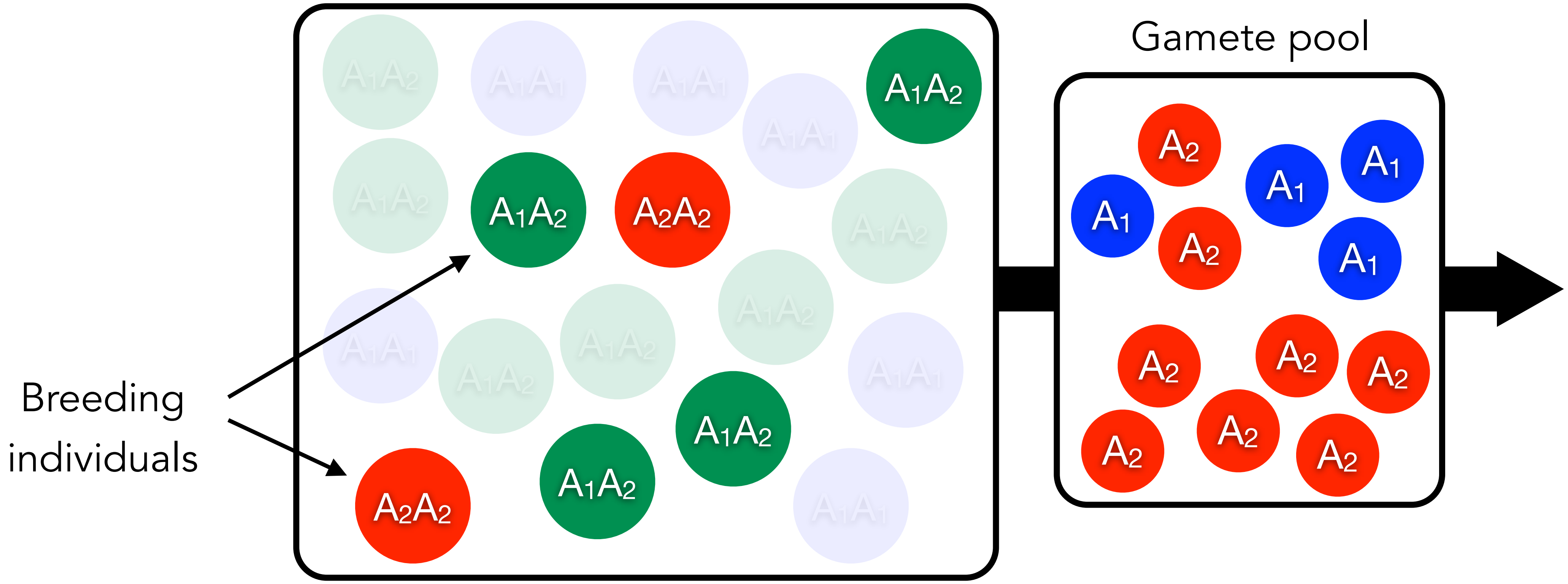
- Mating is random.
- There are equal numbers of males and females, all of whom are able to reproduce.
- All individuals are equally likely to produce offspring, and the number of offspring that each produces varies no more than expected by chance.
- The number of breeding individuals is constant from one generation to the next.

Deviations will decrease the effective population size.

PopGen ▷ Effective Population Size

Natural populations have many **different structures and breeding systems**, which have different genetic consequences. Therefore, the **effective size of a population (N_e)** is usually smaller than the census size (N_c), or the number of **potential breeders**, because the real population differs in structure from the assumptions of the idealised population in terms of **sex ratio**, **family size distribution**, **constancy of numbers in successive generations (population size)**, and having **overlapping** rather than discrete **generations**.

N_c = 18 but N_e = 6



Effective Population Size

$$\frac{N_e}{N_c} \Rightarrow 10\%$$

N>100
unequal sex-ration
variance in reproductive success

Frankham (1995)

$$\frac{N_e}{N_c} \Rightarrow 15\%$$

Palstra & Ruzzante (2008)

$$\frac{N_e}{N_c} \Rightarrow 20\%$$

Waples (2002)

PopGen ▷ Effective Population Size

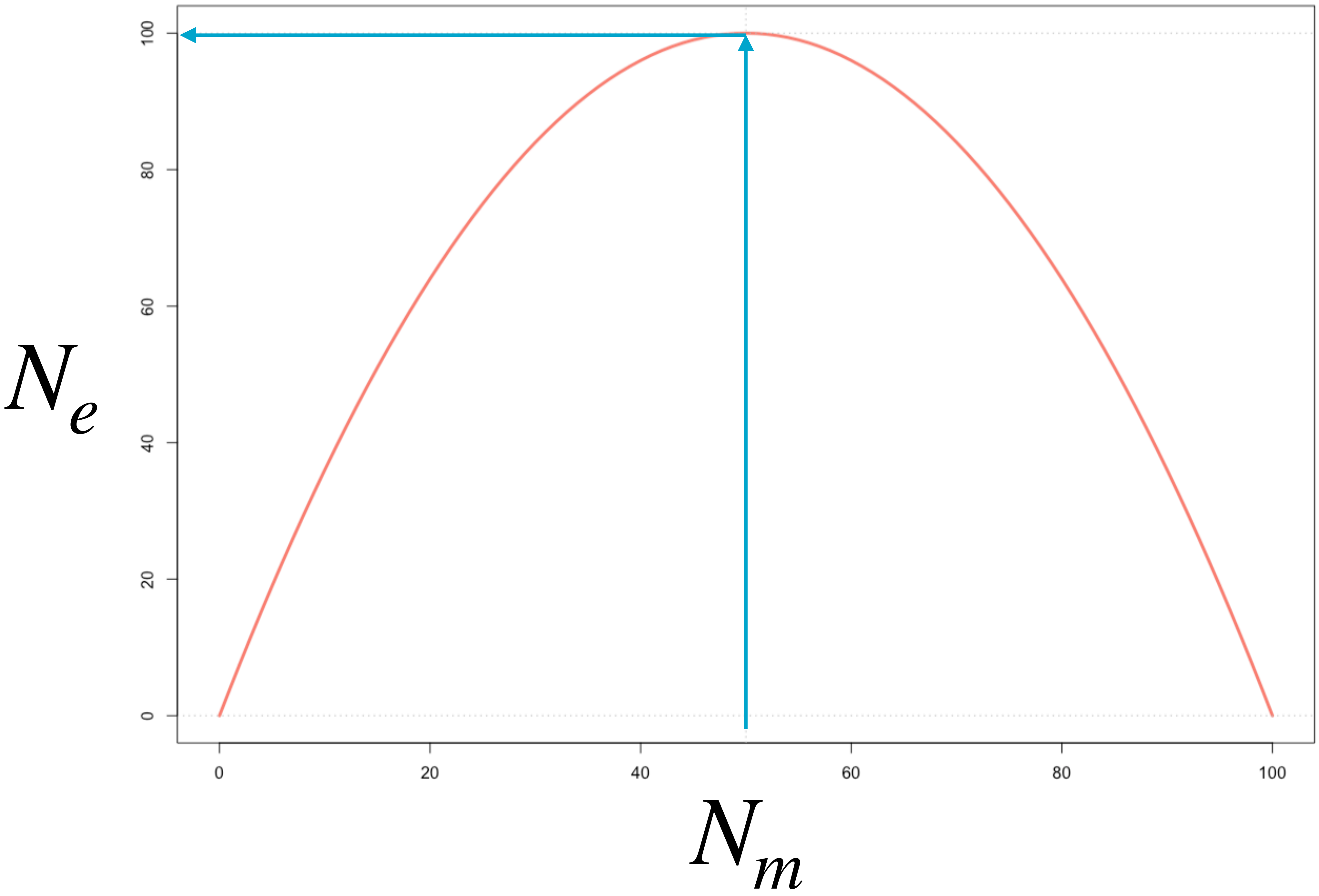
Unequal sex ratios reduce the effective size of the population towards the number in the sex with the fewest breeding individuals.

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

A skewed sex ratio will not have a large effect on the Ne/N ratio unless there is a large excess of one sex over the other.

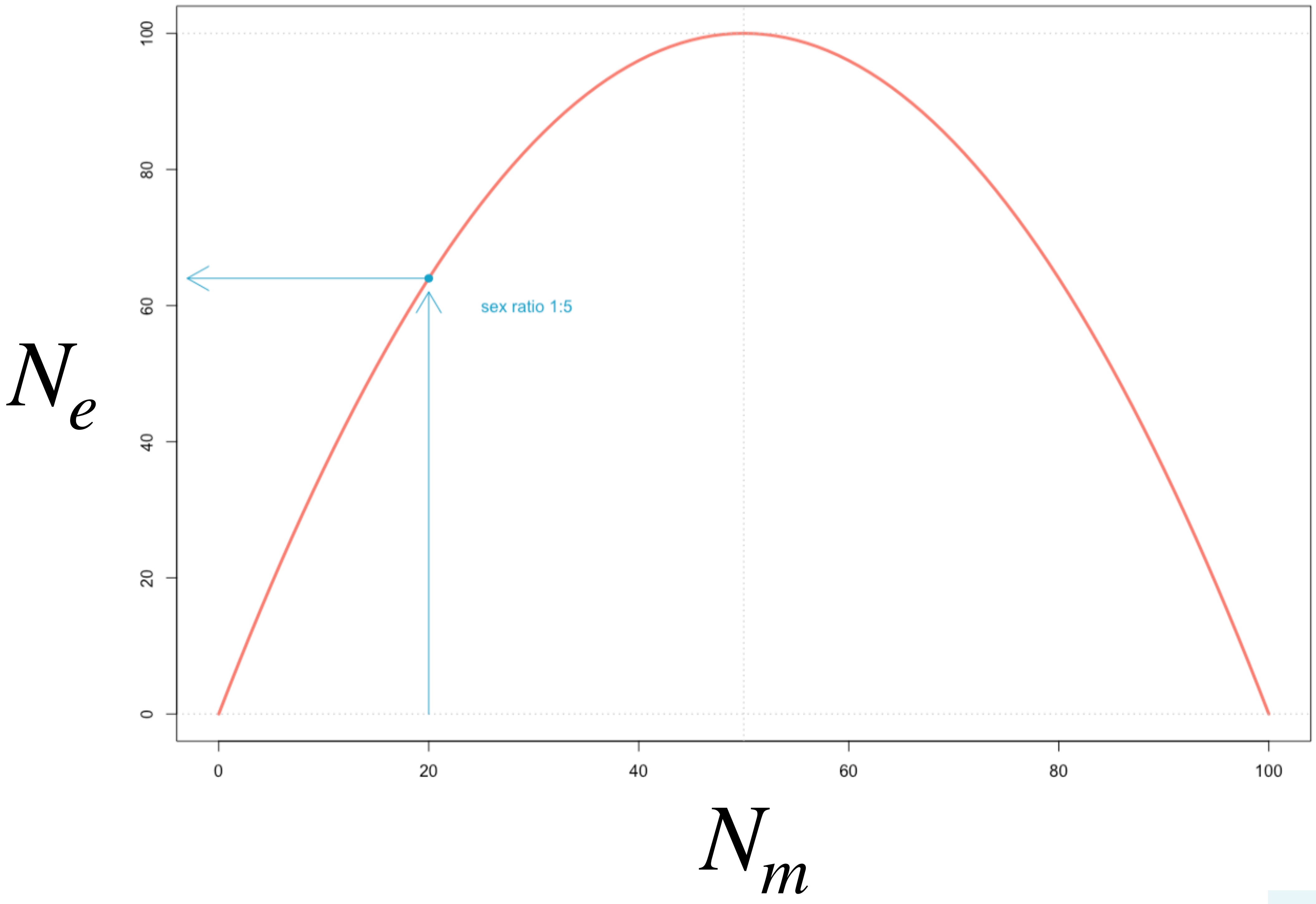
PopGen ▷ Effective Population Size

$$N_m = N_f = 50 \rightarrow N_c = N_e = N_M + N_F = 100$$



Effective Population Size

$$N_m = 20; N_f = 80 \rightarrow N_e = 64$$



PopGen ▷ Effective Population Size

```
Ne <- function(Nm=100-Nf, Nf=100-Nm){
  # Enter Nm or Nf [%]
  Ne <- (4*Nm*Nf)/(Nm+Nf)
  #print(Ne)
}
plot(Ne, xlim = range(0,100),
     col = "salmon",
     lwd = 3,
     ylab = "Ne",
     xlab = "Nm"
)
abline(h   = c(0, 100),
       v   = c(50),
       lty = 3,
       col = "gray")
points(20, 64, col = "#00A5CC", pch = 19)
text(25, 60, "sex ratio 1:5", adj = 0,  col = "#00A5CC")
arrows(20, 0, 20, 62, col = "#00A5CC")
arrows(20, 64, -3, 64, col = "#00A5CC")
```

PopGen ▷ Effective Population Size

```
# Input: %-Male
Ne <- function(x=0.5) {
  (4* x * (1 - x))/( x + (1 - x))
}

#plot equation
ggplot2::ggplot(data.frame(x=c(seq(0,1,0.01))), aes(x=x)) +
  stat_function(fun=Ne, colour = "red") +
  xlab("%Male") +
  ylab("%Ne")
```

PopGen ▷ Effective Population Size

$$N_m = N_f = 1$$

$$N_e = \frac{4N_mN_f}{N_m+N_f} = \frac{4}{2} = 2$$

Effective Population Size

$$N_m = N_f = N$$

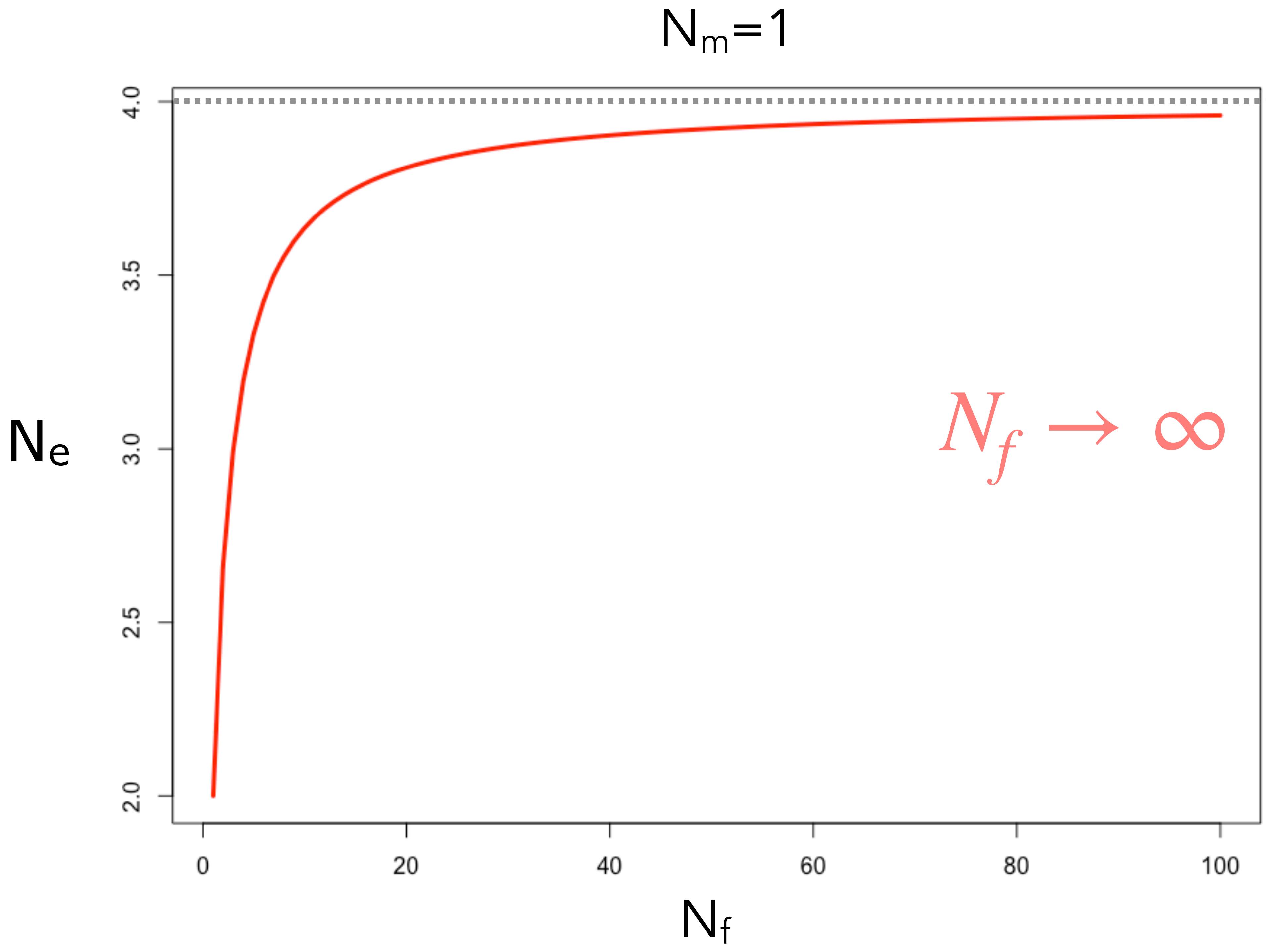
$$N_e = \frac{4N_m N_f}{N_m + N_f} = \frac{4N^2}{2N} = 2N$$

PopGen ▷ Effective Population Size

$$N_m \ll N_f$$

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

PopGen ▷ Effective Population Size



$N_e \approx 4$

$N_f \to \infty$

PopGen ▷ Effective Population Size

```
curve((4 * x) / (x + 1),
      from = 1,
      to = 100,
      xlab = "Nfemale",
      ylab = "Ne",
      main = "Nm = 1",
      type = "l",
      col = "red",
      lwd = 1
)
```

Effective Population Size

$$N_m \gg N_f$$

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

PopGen ▷ Effective Population Size

$$N_m \gg N_f$$

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

PopGen ▷ Effective Population Size

For alleles at an **X-linked** gene, or for alleles in a haplo-diploid organism, the effective population size is slightly different from that for autosomal genes because, for example, females contain two-thirds of the alleles and males contain one-third of the alleles.

$$N_e = \frac{9N_mN_f}{4N_m + 2N_f}$$

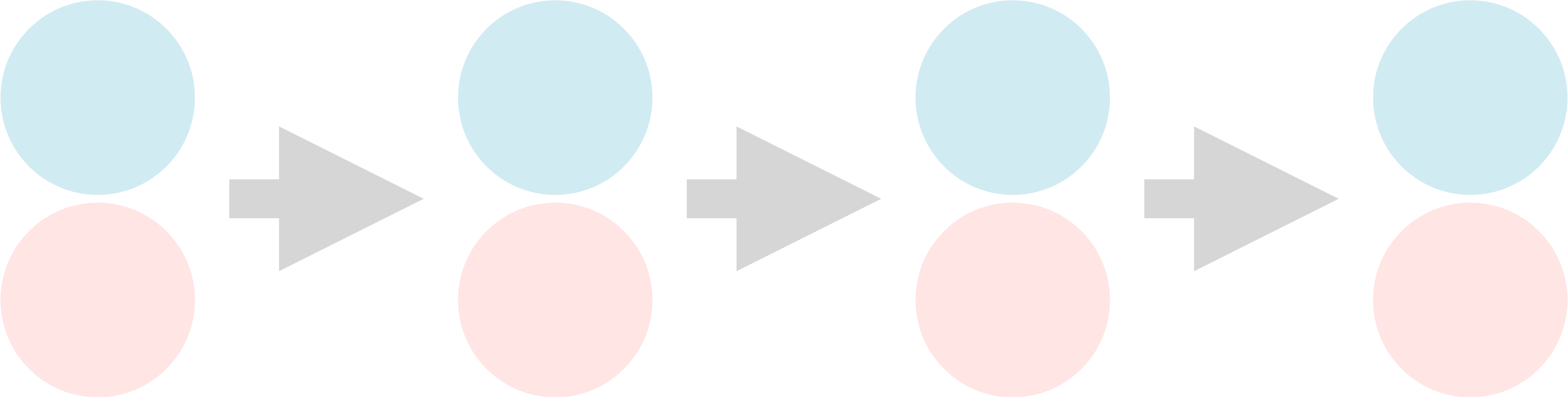
PopGen ▷ Effective Population Size

$$N_m = N_F = 1$$

$$\boxed{N_e = \frac{9N_mN_f}{4N_m+2N_f} = \frac{3}{2}}$$

PopGen ▷ Effective Population Size

Number of offsprings (k) and variation (v_k) are stable over time in a population at Hardy-Weinberg.



Number of offspring (k): **2**

Variance V_k: **2**

Hardy-Weinberg Population → Probability Distributions

Probability Distributions

- Binomial Distribution
- Normal Distribution
- **Poisson Distribution**

- named after the French mathematician Simeon Denis Poisson (1781-1840)
- probability model in biology and medicine
- **the mean and the variance are equal**

PopGen ▷ Effective Population Size

The total fertility rate (TFR) of a population is the average number of children that would be born to a female during her lifetime. Family sizes (**number of offspring**) in wild populations typically show greater variation than the random (Poisson) distribution assumed for the idealised population.

PopGen ▷ Effective Population Size

In general, to account for the variance in the number of offspring, and the population changes in size such that k≠2, the effective population size is approximately:

$$N_e = \frac{N\bar{k}-2}{\bar{k}-1+\frac{V_k}{\bar{k}}}$$

k: number of progeny

V_k: variance in family size

where N is the number of adults in the previous generation, k is the number of offspring (gametes) per parent, and Vk is the variance in family size (variance in the number of offspring).

PopGen ▷ Effective Population Size

Suppose we have an idealised population with an expected number of **offspring of 2**.

$$k = 2$$

Furthermore, in these idealised populations, the expected number of offspring would follow a **Poisson distribution**, where the mean is equal to the variance.

$$\text{Poisson distribution: } k = v_k = 2$$

PopGen ▷ Effective Population Size

$k = 2$

$$N_e = \frac{N\bar{k} - 2}{\bar{k} - 1 + \frac{V_k}{\bar{k}}} = \frac{N2 - 2}{2 - 1 + \frac{V_k}{2}} = \frac{4N - 4}{2 + V_K}$$

Poisson distribution: $k = v_k = 2$

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

PopGen ▷ Effective Population Size

Another commonly used equation for estimating effective population size (N_e) considering variance in the number of progeny for both males and females can be estimated with the following equation:

$$N_e = \frac{4N_m N_f}{V_K(N_m + N_f)}$$

V_{kf}: variance in number of progeny of females

V_{km}: variance in number of progeny of males

PopGen ▷ Effective Population Size

$$N_e = \frac{4N_mN_f}{V_K(N_m + N_f)}$$

if the population is in Hardy-Weinberg equilibrium and mating is random with no other forces affecting reproductive success, then we can typically assume that the variance in the number of progeny for is close to 2.

$$V_k = 2$$

$$N_e = \frac{4N_mN_f}{2(N_m + N_f)} = \frac{2N_mN_f}{N_m + N_f}$$

$$\text{Harmonic Mean} = \frac{2ab}{a + b}$$

If $N_m = N_f$, then $N_m + N_f = N$, and the harmonic mean of N_m and N_f equals N. However, if N_m and N_f differ, N_e will be less than N, reflecting reduce genetic diversity due to uneven sex ration.

PopGen ▷ Effective Population Size

Many species are **not monogamous** and therefore have different numbers of offspring. If males mate with more than one female, Vk is likely to be different for females and males.

$$V_{kf} \neq V_{km}$$

PopGen ▷ Effective Population Size

A commonly used equation for estimating effective population size (N_e) that accounts for variance in the number of progeny among both females and males is:

$$N_e = \frac{8N}{V_{k_f} + V_{k_m} + 4}$$

V_{kf}: variance in number of progeny of females

V_{km}: variance in number of progeny of males

N: total number of breeding individuals (males + females)

This formula takes into account the contribution of reproductive variance for both sexes, which can have a significant effect on effective population size. Higher variance in reproductive success (offspring per individual) typically leads to a lower effective population size compared to the actual number of breeding individuals, as unequal reproductive contributions reduce genetic diversity over generations.

PopGen ▷ Effective Population Size

if the population is in Hardy-Weinberg equilibrium and mating is random with no other forces affecting reproductive success, then we can typically assume that the variance in the number of progeny for both males (V_{km}) and females (V_{kf}) is close to 2.

$$V_{k_f} = V_{k_m} = 2$$

$$N_e = \frac{8N}{V_{kf} + V_{km} + 4} = \frac{8N}{2+2+4} = N$$

PopGen ▷ Effective Population Size

$$V_{kf} < V_{km}$$

Males can often increase their reproductive success by mating with multiple females, leading to a wider variance in the number of offspring. Some males may father many offspring, while others may not reproduce at all.

Females, however, typically have a more limited number of offspring due to the physical constraints of gestation, egg production, or parental care. This limitation generally results in a narrower range of reproductive output among females.

This concept is consistent with **Bateman's principle**, which suggests that reproductive variance is often higher in males because females invest more in offspring production and are therefore more selective about mates, while males can benefit from mating with multiple partners.

Bateman, A. J. (1948). Intra-sexual selection in Drosophila. Heredity, 2(3), 349–368.

PopGen ▷ Effective Population Size

In humans, reproductive variance between the sexes can also be significant. For example, according to the Guinness Book of World Records, the highest recorded number of children born to a single woman is 69. By contrast, the last Sharif emperor of Morocco is estimated to have fathered around 1,400 children. Similarly, modern practices such as sperm donation can lead to cases where a single man has numerous offspring.

* http://www.guinnessworldrecords.com/world-records/3000/most-prolific-mother-ever

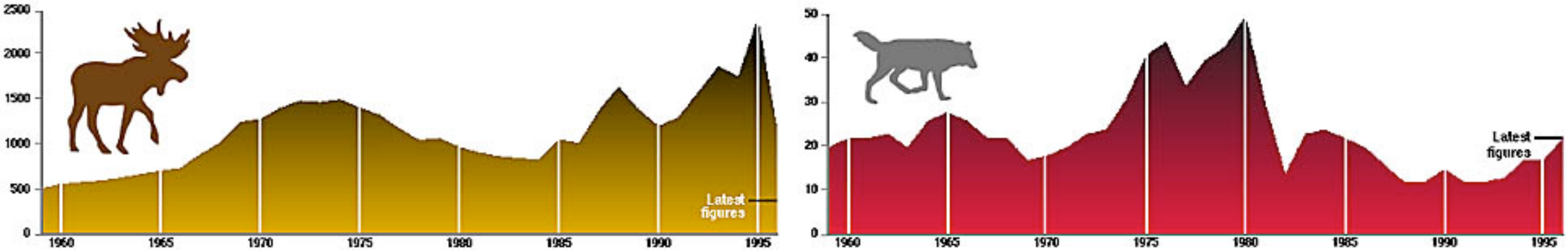
PopGen ▷ Effective Population Size

A 2003 study (Zergal et al) suggested that the DNA of the Mongolian conqueror Genghis Khan is present in some 16 million men alive today. But he's not the only man whose reproductive activities still have a significant genetic impact centuries later. A more recent study (Balaresque et al. 2015) found a handful of other men who founded prolific lineages.

Zergal et al. (2003) Am. J. Hum. Genet. (72) 717-721.
Balaresque et al. (2015) Eur. J. Hum. Genet. (23) 1413–1422.

PopGen ▷ Effective Population Size

Wild populations fluctuate in numbers as a result of changing climatic conditions, disease, epidemics, varying hunting pressure and many other factors.



Moose and wolf population fluctuations on Isle Royale National Park in Lake Superior depend on factors like weather, disease and possibly genetic problems.

Fluctuations in population size over generations reduce N_e below the average number of adults. When calculating the effective population size across multiple generations or time periods, the harmonic mean accounts for fluctuations in population size. This is given by:

$$\frac{1}{N_e} = \frac{1}{t}\left(\frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{N_3} + \ldots + \frac{1}{N_t}\right)$$

Here, *t* is the number of time periods, and N_1, N_2,…,N_t represent the population size at each time point. This approach reflects that low population sizes in any time period disproportionately reduce the overall N_e.

PopGen ▷ Effective Population Size

The **harmonic mean** is a way of calculating an average that gives more weight to smaller values. It's often used in situations where values are not evenly distributed and small values are particularly influential. In general, it's useful when averaging rates, proportions or populations over time or groups, where small numbers have a greater effect on the result.

The harmonic mean is particularly sensitive to smaller values. A small value in the dataset will pull the harmonic mean down more than it would affect an arithmetic mean.

For any set of positive values, the harmonic mean will always be less than or equal to the arithmetic mean.

PopGen ▷ Effective Population Size

Heterozygosity after one generation:

$$H_1 = \left(1 - \frac{1}{2N}\right) H_0$$

Heterozygosity after t generations:

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

PopGen ▷ Effective Population Size

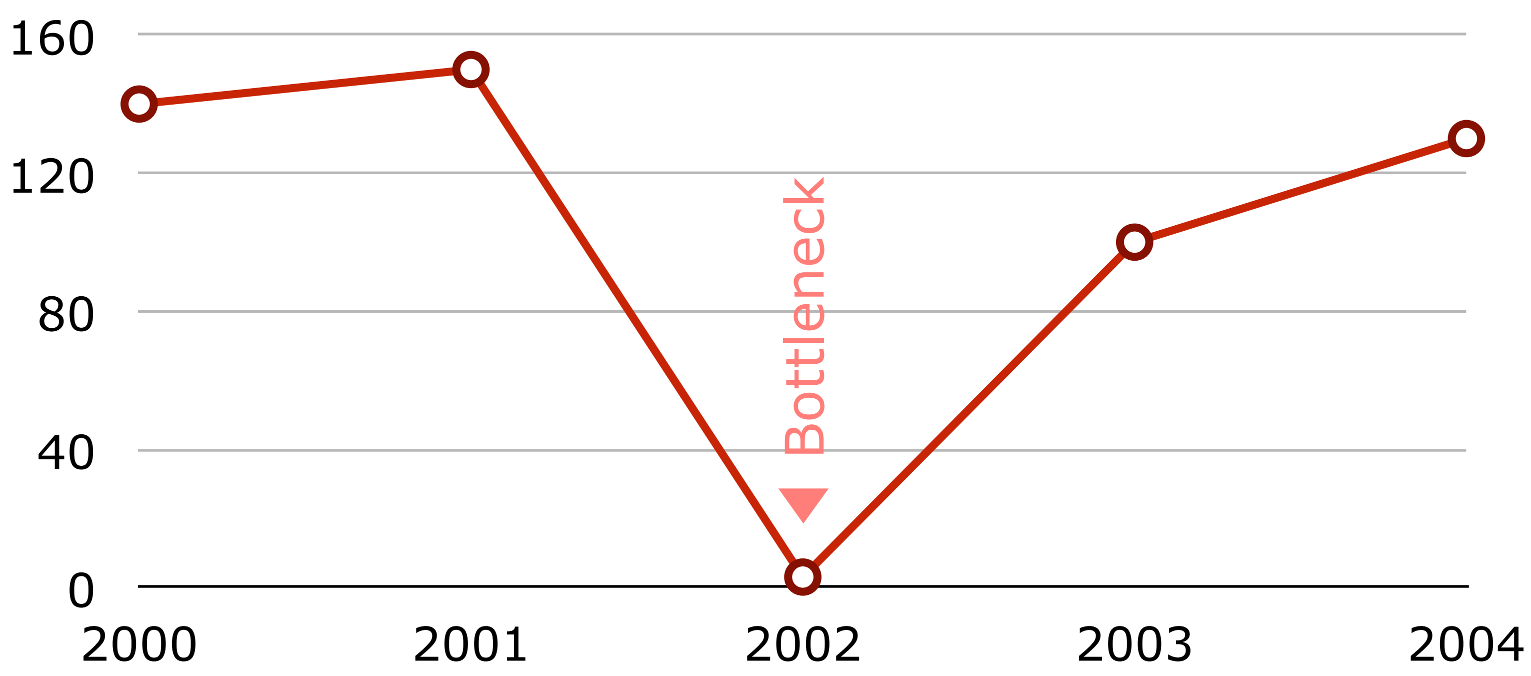
Year	N
2000	140
2004	130

$$\text{t} \rightarrow 1\text{year} \rightarrow 1\text{generation}$$

$$h_t = \left(1 - \frac{1}{2N}\right)^t h_0 \xrightarrow[N=130]{t=5} \left(1 - \frac{1}{2*130}\right)^5 h_0 = 0.98 h_0$$

↳ The expect lose of the heterozygosity is ~**2%**.

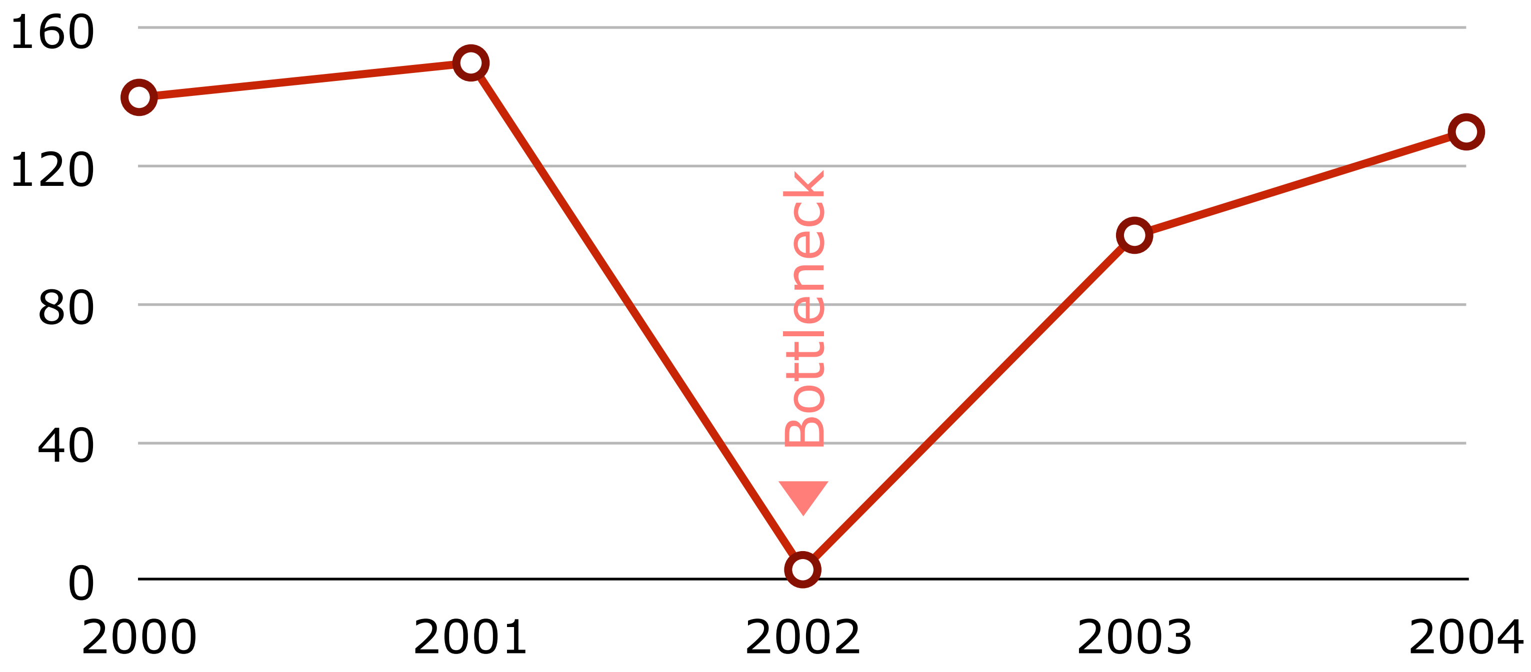
PopGen ▷ Effective Population Size



Year	N
2000	140
2001	120
2002	**3**
2003	100
2004	130

$$\frac{1}{N_e} = \tfrac{1}{5}\left[\frac{1}{140} + \frac{1}{120} + \frac{1}{3} + \frac{1}{100} + \frac{1}{130}\right] = 0.0733 \Rightarrow N_e = 13.6$$

PopGen ▷ Effective Population Size

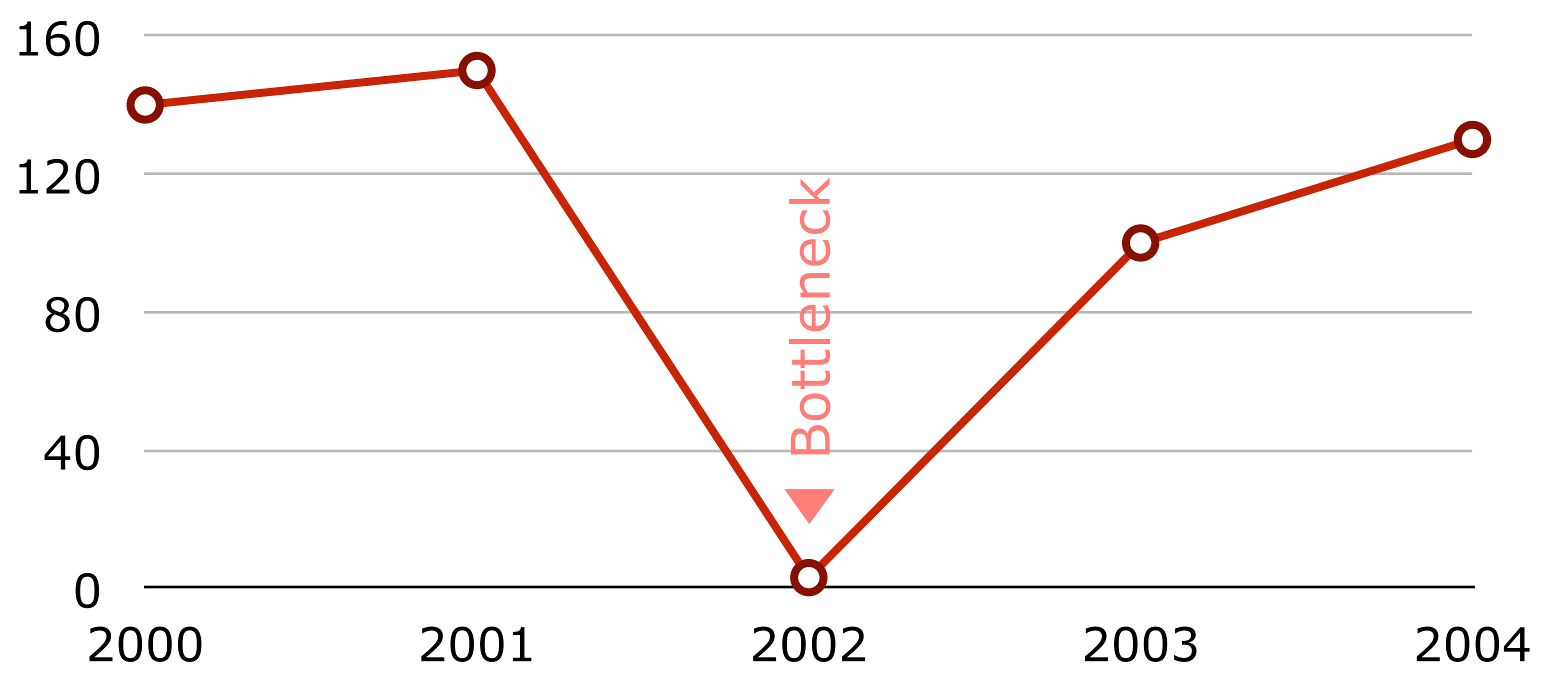


Year	N
2000	140
2001	120
2002	**3**
2003	100
2004	130

$$h_t = \left(1 - \frac{1}{2N_e}\right)^t h_0 \xrightarrow[N=13.6]{t=5} \left(1 - \frac{1}{2*13.6}\right)^5 h_0 = 0.83h_0$$

↳ Taking into account fluctuations in population size over time, the loss of heterozygosity is **17%**.

PopGen ▷ Effective Population Size



Year	N
2000	140
2001	120
2002	**3**
2003	100
2004	130

The probability of an allele being lost after a bottleneck is: $(1-p)^{2N}$

$$p = 0.50; N_e = 3 \Rightarrow \sim 1.6\%$$

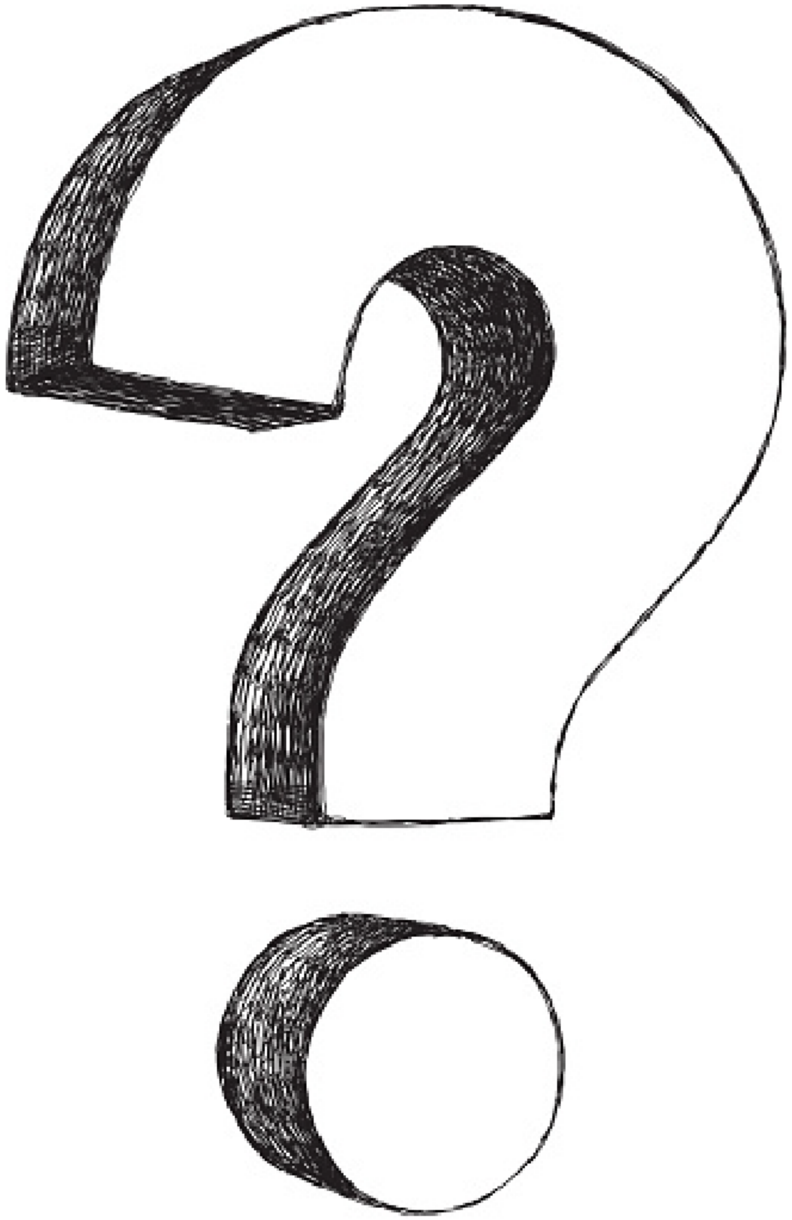
$$p = 0.25; N_e = 3 \Rightarrow \sim 17.8\%$$

The rarer an allele is, the more likely it is to be lost in a bottleneck.

PopGen ▷ Effective Population Size

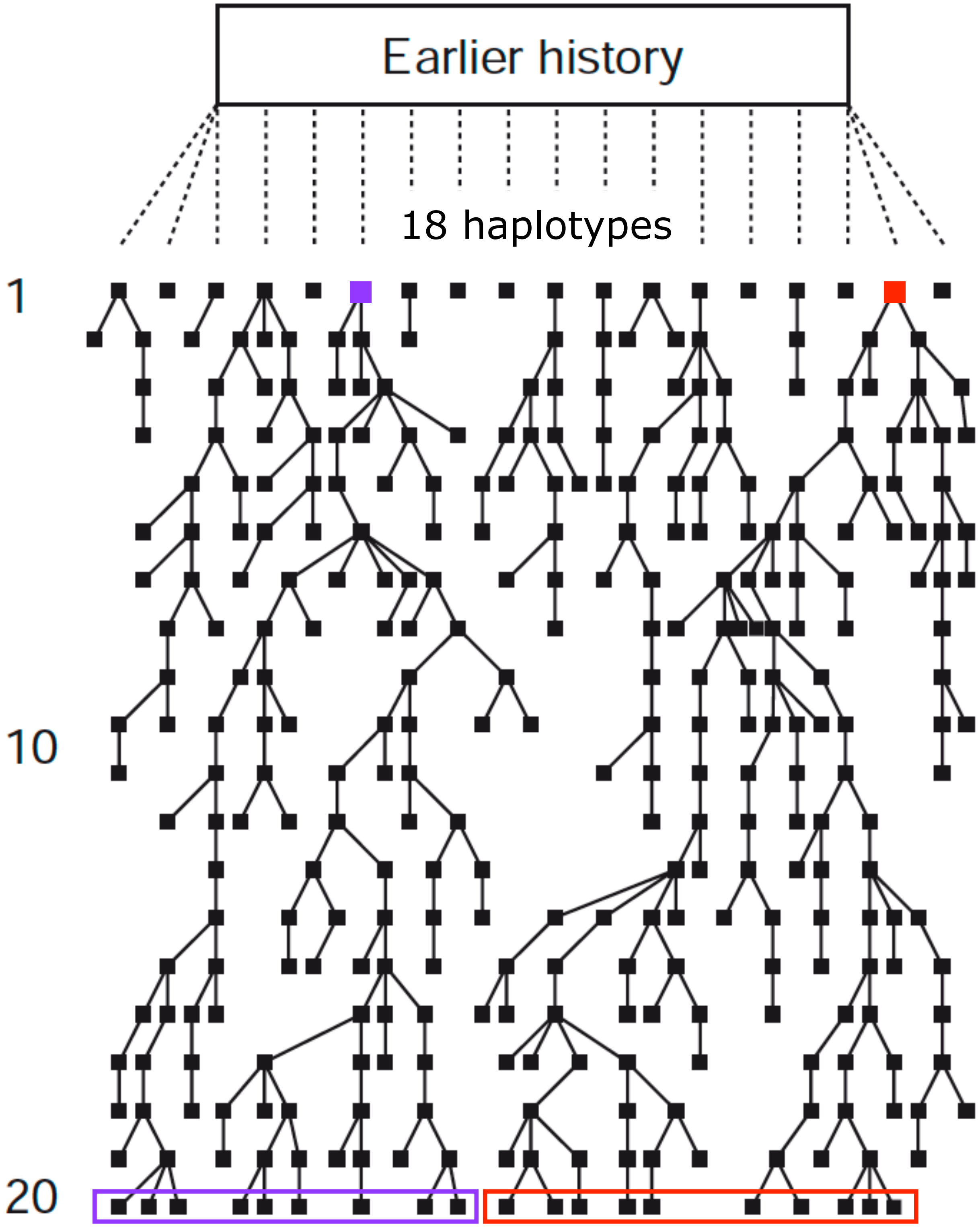
A species that loses genetic variation to drift (e.g. because its census population has gone through a severe bottleneck) will find it very difficult to recover the lost variation, because Ne is most sensitive to the smallest population sizes over time. In fact, even if the census size of the population can be increased (perhaps through captive breeding efforts), genetic variation may continue to decrease because Ne still reflects the recent bottleneck.

PopGen ▷ Effective Population Size



MORE
THINGS
CONSIDERED

PopGen ▷ Effective Population Size



Allendorf, Luikart, and Aitken 2013

There is an alternative approach to study the loss of genetic variation in populations. It is possible to trace the transmission of e.g. mtDNA haplotypes over many generations. That is, we can trace the genealogy of the mtDNA genotype of each individual in a population.