Drift When Populations Vary in Size

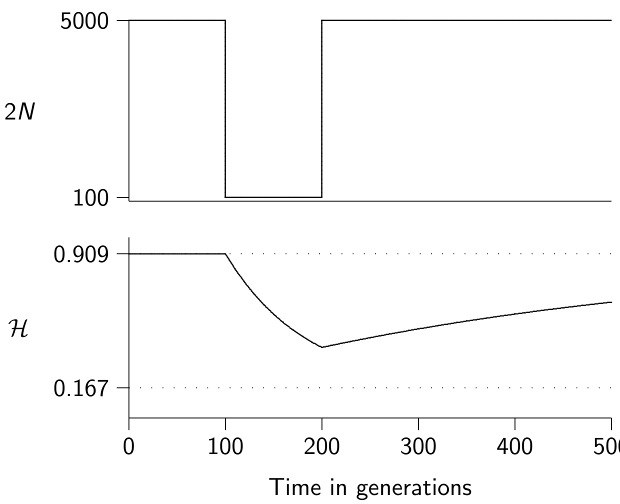
Alan R. Rogers

February 9, 2023

Why is heterozygosity so often lower than we expect?

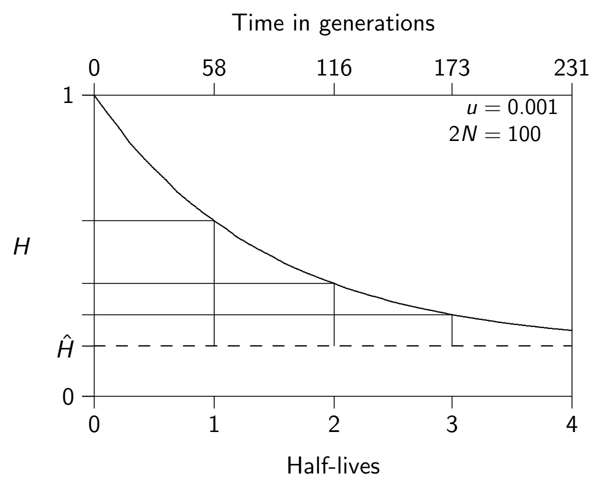
- Urn model assumes N is constant. What if it varies?
- Bottleneck: a temporary reduction in N
- Decline in $\mathcal{H}$ is faster than recovery.
- Effective population size is harmonic mean of N_t
- Harmonic mean is sensitive to small sizes.

A bottleneck in population size



$\mathcal{H}$ declines rapidly, recovers slowly. Why?

What is a half-life?



Why the decline is faster than the recovery

Gene diversity converges toward equilibrium with a half-life of

$$t_h = \frac{\ln 2}{2u + 1/2N}$$

Small $N \Rightarrow$ short half-life.

N has little effect if $2u \gg 1/2N$, i.e. if $\theta \gg 1$.

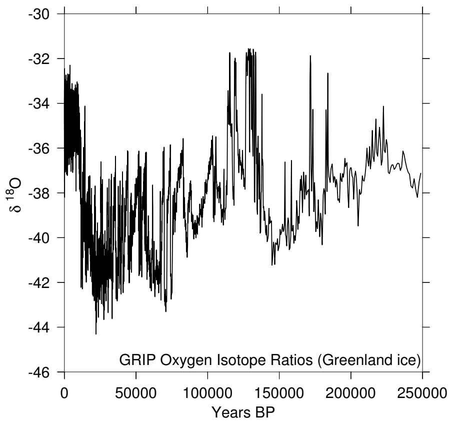
		Half-life of convergence	
$2N$	θ	(gen.)	(years)
∞	∞	347	1,041
10^6	2000.00	346	1,038
10^5	200.00	345	1,035
10^4	20.00	330	990
10^3	2.00	231	693
10^2	0.20	58	174
10	0.02	7	21

(Assumes $u = 0.001$)

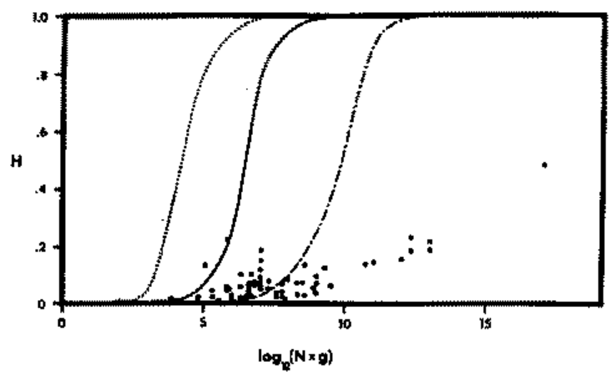
Oscillations in population size

We've been considering a single bottleneck in population size.
What if it is always varying?
To see that this is plausible, consider the record of climate change during the past 250,000 years.

History of global temperature



What did this do to population size?



Heterozygosity, $\mathcal{H}$, versus population size, Ng, where g is ploidy. Solid line: expected curve for neutral alleles; dotted: slightly overdominant; dot-dashed: slightly deleterious.

(Nei & Graur 1984)

For many species, $\mathcal{H}$ is much smaller than would be expected on the basis of their population sizes. Could this be a result of population size bottlenecks during the Pleistocene?

Effective population size, N_e

Goal: Find a value of N that makes our idealized population behave like a more complicated one.
Example: In a randomly mating population of constant size, heterozygosity (gene diversity) is equal to

$$\mathcal{H} = \frac{4Nu}{4Nu + 1}$$

What if the population varies in size?

Review: $\mathcal{H}$ in a population of constant size

$$\mathcal{H}_1 = \mathcal{H}_0 \left(1 - \frac{1}{2N}\right)$$

Another generation

$$\begin{aligned}\mathcal{H}_1 &= \mathcal{H}_0 \left(1 - \frac{1}{2N}\right) \\ \mathcal{H}_2 &= \mathcal{H}_1 \left(1 - \frac{1}{2N}\right)\end{aligned}$$

$$\begin{aligned}\mathcal{H}_1 &= \mathcal{H}_0\left(1-\frac{1}{2N}\right)\\ \mathcal{H}_2 &= \mathcal{H}_1\left(1-\frac{1}{2N}\right)\\ &= \mathcal{H}_0\left(1-\frac{1}{2N}\right)^2\end{aligned}$$

General form

$$\begin{aligned}\mathcal{H}_t &= \mathcal{H}_0\left(1-\frac{1}{2N}\right)^t\\ &\approx \mathcal{H}_0\exp[-t/2N]\end{aligned}$$

Note approximation: $1-x\approx e^{-x}$ when x is small.

$\mathcal{H}$ in a population of varying size

$$\mathcal{H}_1 = \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)$$

where N_0 is population size in generation 0.

Another generation

$$\begin{aligned}\mathcal{H}_1 &= \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)\\ \mathcal{H}_2 &= \mathcal{H}_1\left(1-\frac{1}{2N_1}\right)\end{aligned}$$

Two generations

$$\mathcal{H}_2 = \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)\left(1-\frac{1}{2N_1}\right)$$

Two generations again

$$\begin{aligned}\mathcal{H}_2 &= \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)\left(1-\frac{1}{2N_1}\right)\\ &= \mathcal{H}_0\prod_{i=0}^{1}\left(1-\frac{1}{2N_i}\right)\end{aligned}$$

where $\prod$ is the product operator.

General form

$$\mathcal{H}_t = \mathcal{H}_0 \prod_{i=0}^{t} \left(1 - \frac{1}{2N_i}\right)$$
$$\approx \mathcal{H}_0 \exp\left[-\sum_{i=0}^{t-1} \frac{1}{2N_i}\right]$$

Compare results for fixed and varying N

$$\mathcal{H}_t \approx \underbrace{\mathcal{H}_0 \exp[-t/2N_e]}_{\text{Fixed } N} = \underbrace{\mathcal{H}_0 \exp\left[-\sum_{i=0}^{t-1} \frac{1}{2N_i}\right]}_{\text{Varying } N}$$

- N_e is called effective population size.
- It is the constant population size that makes the two sides equal.

The two sides are equal when

$$1/N_e = \frac{1}{t}\sum_{i=0}^{t-1} \frac{1}{N_i}$$

The effective population size, N_e, is the "harmonic mean" of $N_0, N_1, \ldots, N_{t-1}$.

What is N_e good for?

In a population of varying size, average heterozygosity at neutral loci is

$$\mathcal{H} = \frac{4N_e u}{4N_e u + 1}$$

where N_e is the effective population size.

Example

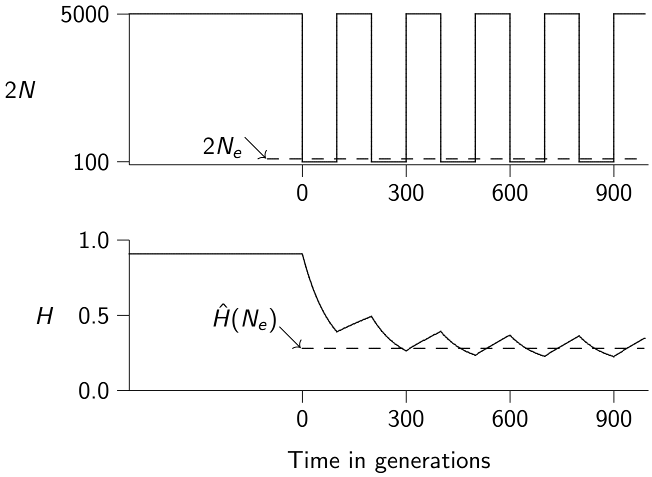
- What is the arithmetic mean of 1, 50, and 100?
- What is the harmonic mean?

Answer

- Arithmetic mean: $(1 + 50 + 100)/3 = 50.3333$.
- Harmonic mean: $1/((1 + 1/50 + 1/100)/3) = 2.9126$.

Harmonic mean is *much* smaller than arithmetic mean.

Approach toward equilibrium when size varies

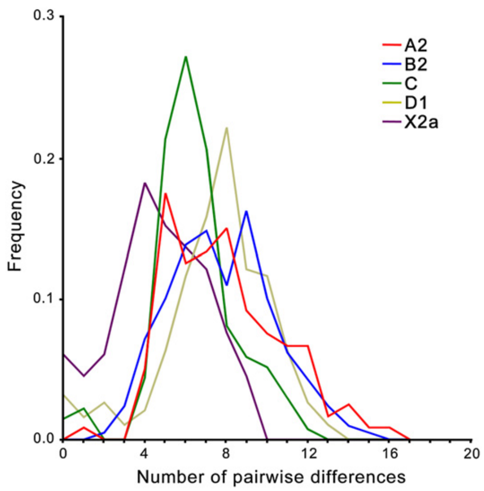


Genetics and the History of Population Size

Alan R. Rogers

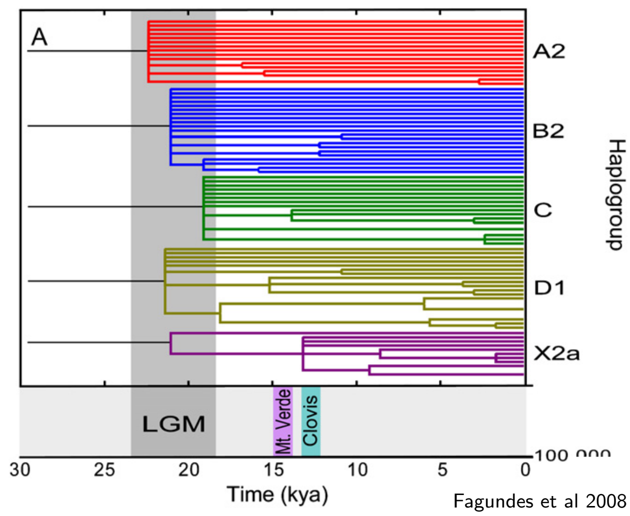
February 9, 2023

Mismatch Distributions of Amerindian mtDNA Haplogroups



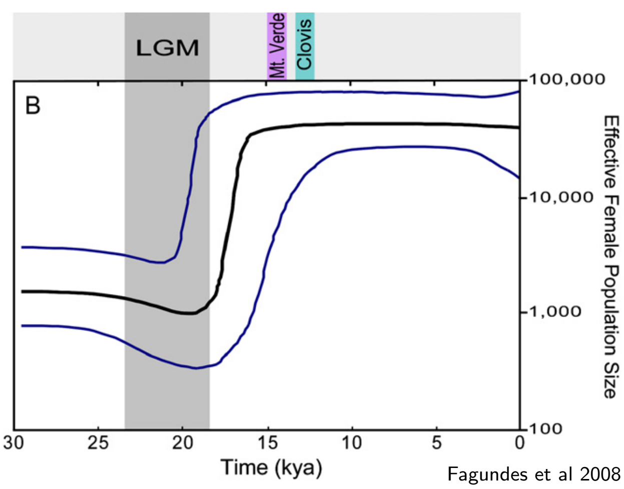
Fagundes et al 2008

Genealogies of Amerindian mtDNA Haplogroups



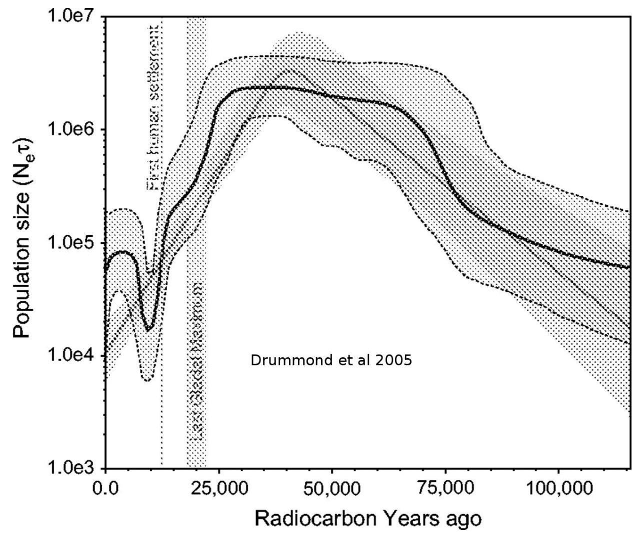
Fagundes et al 2008

Estimated Size of Amerindian Population



Fagundes et al 2008

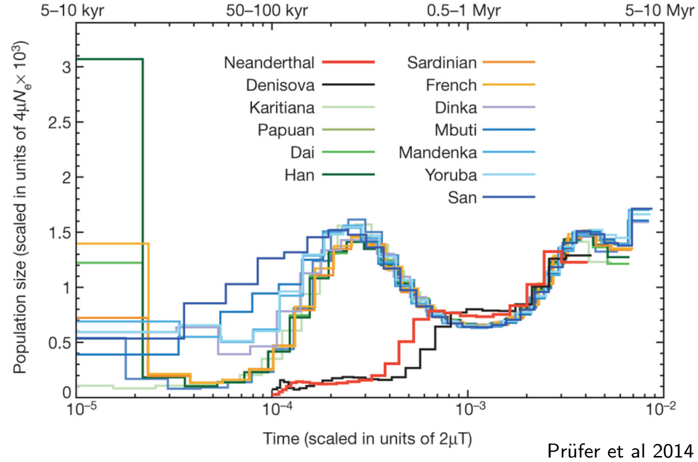
Estimated Size of Bison Population



What about the nuclear genome

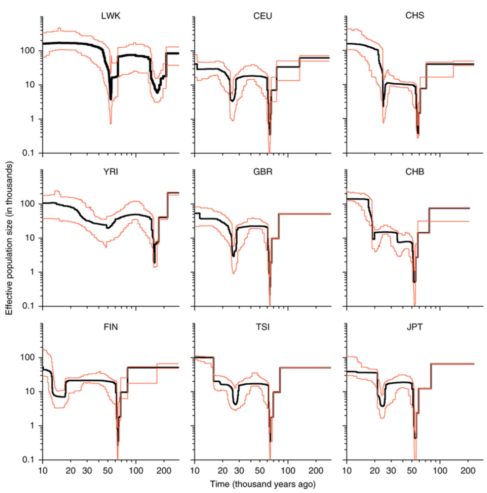
- Huge amounts of data.
- Recombination makes previous methods unusable.

PSMC: deep history from a single diploid genome



Prüfer et al 2014

Accurate back to 2 mya. Not for last 20,000 years.



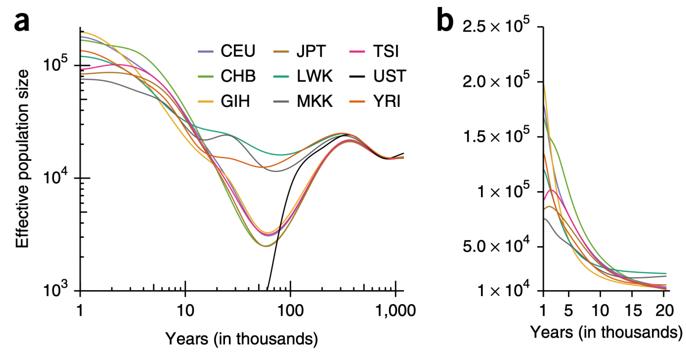
Stairway plot

Uses site frequency spectrum.

Accomodates large samples. Can study last 20,000 years.

Liu & Fu 2015

SMC++



8 modern populations and Ust'-Ishim (45-kya modern Siberian). Log scale on left, arithmetic on right. Combines advantages of PSMC and spectrum. Large samples or small; accurate across both recent and deep scales of time. (Terhorst et al. 2017)