

How Mutation Affects the Decay of Heterozygosity

Alan R. Rogers

January 25, 2024

Two models of mutation

The mutation rate is u per gamete per generation.

Infinite alleles Each mutant is an allele never seen before.

K alleles When allele i mutates, the mutant is equally likely to be any allele other than i . There are $K - 1$ possibilities, each with probability $1/(K - 1)$.

We'll focus on the model of infinite alleles.

Model of infinite alleles

- ▶ Each mutation creates a unique mutation, which has never been seen before.
- ▶ Two identical gene copies remain identical in next generation only if neither mutates.
- ▶ Probability of this is $(1 - u)^2$, where u is the mutation rate.

How drift affects gene identity

Without mutation

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) \mathcal{G}$$

With mutation

$$\mathcal{G}' = (1 - u)^2 \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) \mathcal{G} \right]$$

Approximations

$$\begin{aligned}(1-u)^2 &\approx 1-2u \\ \frac{1-2u}{2N} &\approx \frac{1}{2N}\end{aligned}$$

Numerical example: $(1 - u)^2 \approx 1 - 2u$

u	$(1 - u)^2$	$1 - 2u$
0.100000	0.810000	0.800000
0.010000	0.980100	0.980000
0.001000	0.998001	0.998000
0.000100	0.999800	0.999800
0.000010	0.999980	0.999980
0.000001	0.999998	0.999998

Numerical example: $(1 - u)/2N \approx 1/2N$

u	$2N$	$(1 - 2u)/2N$	$1/2N$
0.0001	10	0.0999800	0.10000
0.0001	100	0.0099980	0.01000
0.0001	1000	0.0009998	0.00100
0.0001	10000	0.0001000	0.00010
0.0001	100000	0.0000100	0.00001

Before approximations

$$\mathcal{G}' = (1 - u)^2 \left[\frac{1}{2N} + \left(1 - \frac{1}{2N} \right) \mathcal{G} \right]$$

After

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - 2u - \frac{1}{2N} \right) \mathcal{G}$$

At equilibrium $\mathcal{G}' = \mathcal{G}$, so

$$\begin{aligned}\hat{\mathcal{G}} &= \frac{1}{4Nu + 1} \\ \hat{\mathcal{H}} &= 1 - \hat{\mathcal{G}} = \frac{4Nu}{4Nu + 1}\end{aligned}$$

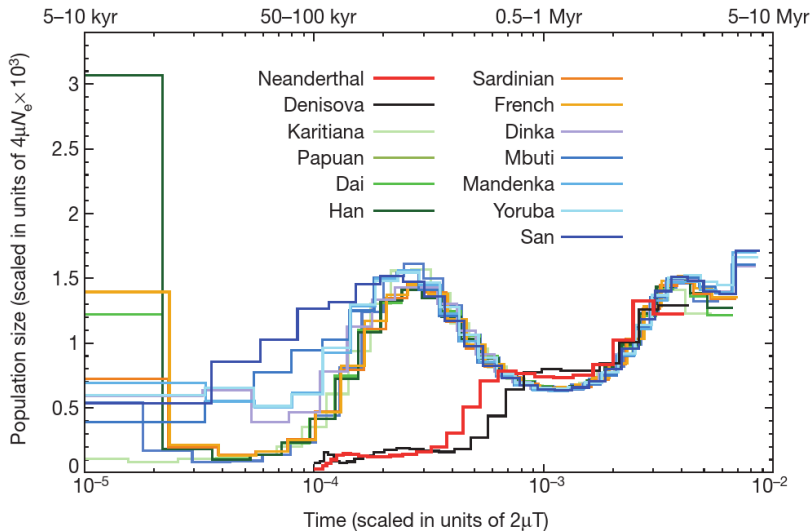
for model of infinite alleles. The “hats” indicate that these are equilibrium values. If $4Nu$ is large, $\hat{\mathcal{H}} \approx 1$.

Model of 2 alleles

$$\hat{\mathcal{H}} = \frac{4Nu}{8Nu + 1}$$

If $4Nu$ is large, $\hat{\mathcal{H}} \approx 1/2$.

History of human population size



Neanderthals had low heterozygosity

Species	Population	Heterozygosity
Neanderthal	El Sidrón	0.000143
	Vindija	0.000127
	Chagyrskaya	0.000147
	Altai	0.000113
Modern	African	0.000507
	European	0.000387
	Asian	0.000358

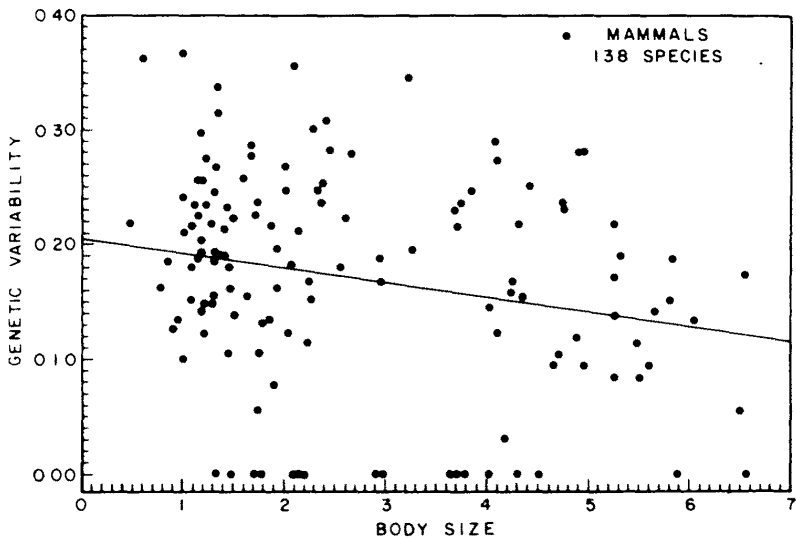
Low heterozygosity \Rightarrow small population.

The magnitude of predicted effects on heterozygosity at a biallelic locus

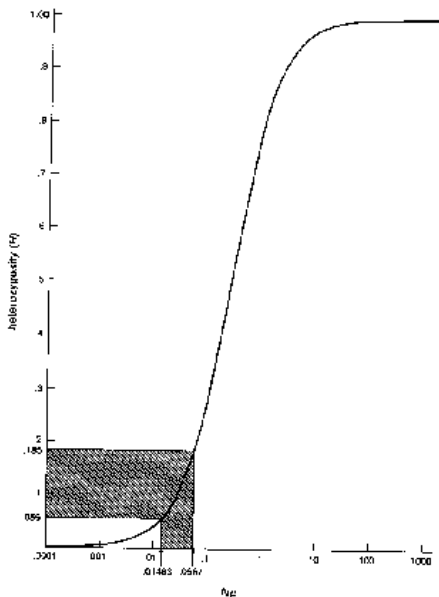
N	H
1,000	0.00004
10,000	0.00040
100,000	0.00400
1,000,000	0.03700
$u = 10^{-8}$	

Populations of different size should differ enormously in heterozygosity.

Is this really true?



Heterozygosity at enzyme polymorphisms vs \log_{10} grams of body wgt (Wooten & Smith 1985). Range: 3 g to 4000 kg.



Variation in H implies implausibly-small variation in $N\mu$.

Lewontin (1970)

Puzzles

- ▶ Why is there so little heterozygosity?
- ▶ Small animals have large populations and should have high heterozygosity. Why don't they?

Much of the rest of this course is about these questions.

A Python program to calculate $\hat{\mathcal{H}}$, using biallelic model

$$\hat{\mathcal{H}} = \frac{4N_u}{8N_u + 1} = \frac{\theta}{2\theta + 1}$$

```
# Expected heterozygosity as a function
# of theta = 4*N*u
def h(theta):
    return(theta/(2*theta + 1.0))

for theta in [0.001, 0.01, 0.1, 1.0, 10, 100]:
    print("%8.3f %8.3f" % (theta, h(theta)))
```

Expected heterozygosity at a biallelic locus

$$\hat{\mathcal{H}} = \frac{\theta}{2\theta + 1}$$

θ	$\hat{\mathcal{H}}$
0.001	0.001
0.010	0.010
0.100	0.083
1.000	0.333
10.000	0.476
100.000	0.498

Expected heterozygosity at a biallelic locus

