neutral evolution $\Rightarrow$ heterozygosity

$$\mathcal{H} = \frac{4N\mu}{1+4N\mu} \equiv \frac{\Theta}{1+\Theta}$$

↑ measure this from gen. seq. populations

$$\Theta = 4N\mu \qquad \approx \Theta$$

↑ can also be measured

measure $\mathcal{H} \Rightarrow$ get $\Theta \sim 1.1 - 2 \times 10^{-3}$ (for single bp in humans)

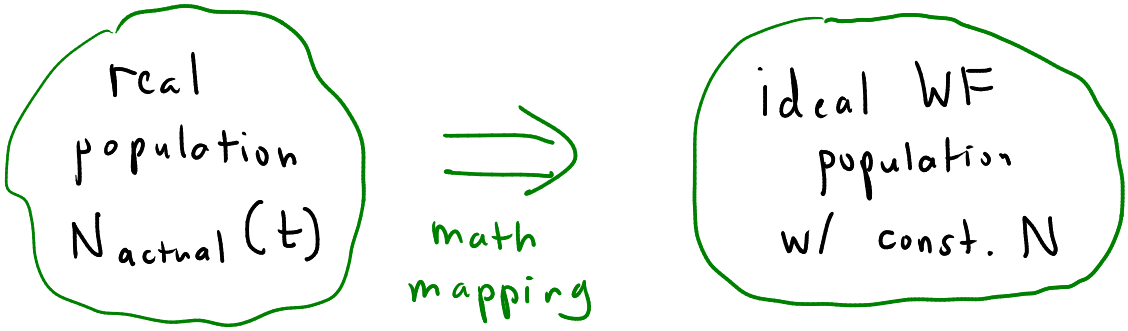
$\mu \sim 2.6 \times 10^{-8}$ / bp / generation

$\Rightarrow N = 1 - 2 \times 10^4$ population # you get from fitting humans to a neutral WF

compare to $N_{actual} \sim 9 \times 10^9$

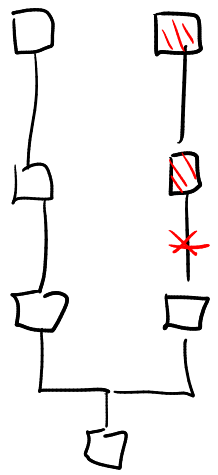
When we fit WF to real population

the N value $\equiv$ effective population size



For small μ, $\mathcal{H} \propto N$ so we can interpret N as a measure of genetic diversity.

Focus on another quantity:



prob. that two lineages of a pair of alleles in curr. generation have not coalesced going back n generations:

$$U_c(n) \simeq \left(1 - \frac{1}{2N}\right)^n$$ in ideal neutral WF model

$t = n t_g$ ↓ gen. time (20 yrs for humans)

$$U_c(t) \simeq \left(1 - \frac{1}{2N}\right)^{\frac{t}{t_g}} \approx \left(e^{-\frac{1}{2N}}\right)^{\frac{t}{t_g}} = \boxed{e^{-\frac{t}{2N t_g}}}$$

$N >> 1$

WF dynamics but w/ time varying $N_a(t)$ ↑ actual

$$\Rightarrow U_c(t) = \left(1 - \frac{1}{2N_a(-t_g)}\right) \cdot \left(1 - \frac{1}{2N_a(-2t_g)}\right)$$

$t = 0$ now
$t = -t_g$ 1 gen. ago
...

$$\cdots \left(1 - \frac{1}{2N_a(-n t_g)}\right)$$

$$\approx \exp\left[-\frac{1}{t_g} \sum_{j=1}^{n} t_g \frac{1}{2N_a(-jt_g)} \right]$$

$$\approx \boxed{\exp\left[-\frac{1}{t_g} \int_0^t dt' \frac{1}{2N_a(t')} \right]}$$

WF w/ time-varying Pop. $N_a(t)$

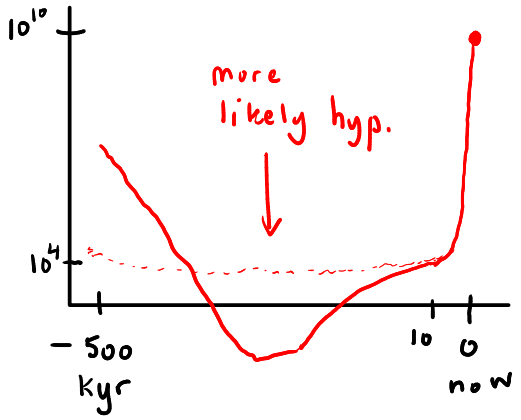
in order for ideal + time-varying model to give same answer for $U_c(t) \Rightarrow$

$$N_{\text{eff. pop. size}} = \left[\underbrace{\frac{1}{t} \int_0^t dt' \frac{1}{N_a(t')}}_{\text{avg. over time}} \right]^{-1}$$

$$\Rightarrow R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

dominated by small populations in past

Sudden drops in past populations $\Rightarrow$ genetic bottlenecks $\Rightarrow$ can lead to small N (low gen. diversity)



How long ago (on average) was coalescence?

prob. that coal. occurred b/t t and $t+dt$ in past:

$$P_c(t)\,dt = U_c(t) - U_c(t+dt) \approx -dt\,\frac{dU_c(t)}{dt}$$

mean coal. time:

$$t_c = \int_0^\infty dt\; t\; P_c(t)$$

$$= -\int_0^\infty dt\; t\,\frac{dU_c}{dt}$$

$$= -\int_0^\infty dt\; t\left(-\frac{1}{2Nt_g}\right)e^{-\frac{t}{2Nt_g}}$$

$$\Rightarrow \quad t_c = 2N\,t_g \quad \Rightarrow \quad t_c = 400{,}000 \text{ yrs}$$

($2N$: 10^4; t_g: 20 yrs)

rough agreem. w/ more det. estimates

mitochondrial Eve: $\sim$ 99 - 152 kyr

Y-chrom. Adam: $\sim$ 120 - 581 kyr

single copy: $\Theta = 2N\mu \Rightarrow N = \begin{cases} N_{fem} \\ N_{male} \end{cases}$