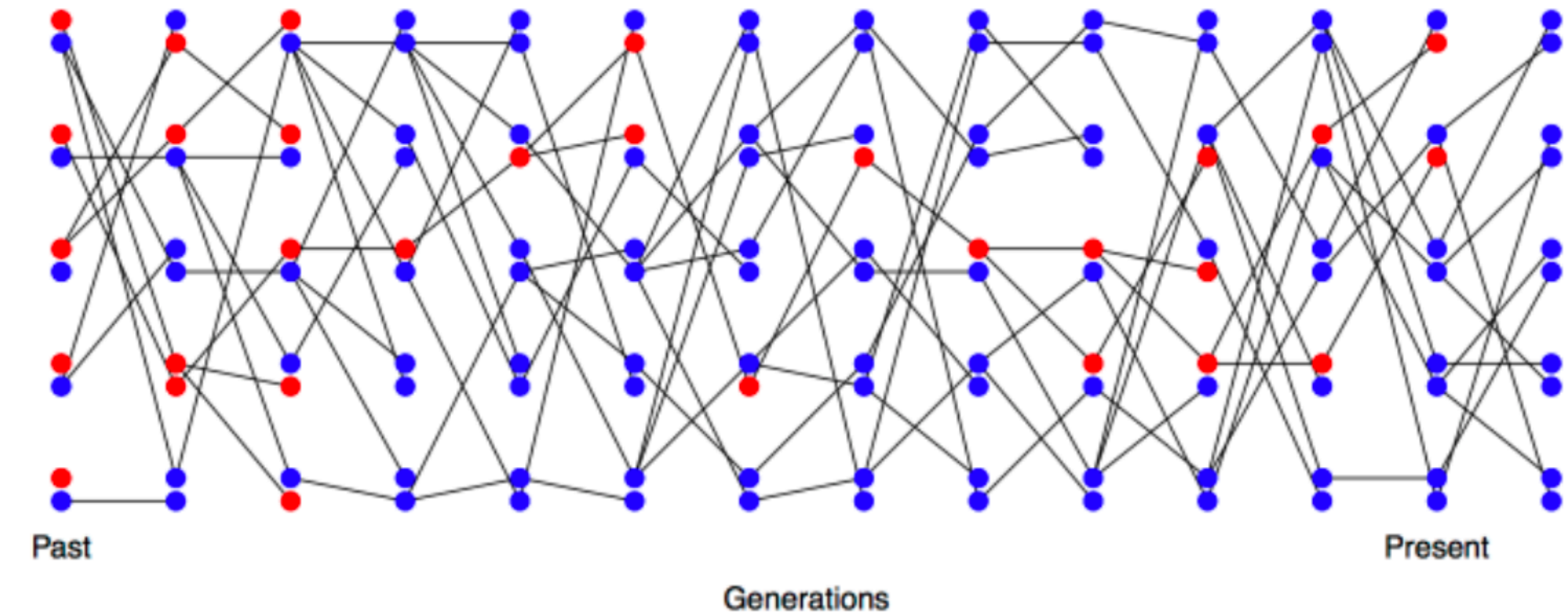


# Lecture 4: Introduction to Genetic Drift

*Population genetic PCB4553/6685*

# Genetic drift

- Genetic drift is the change in the frequency of an existing gene variant (allele) in a population due to random chance.
- Individuals carrying a particular allele can leave more or less offspring in the next generation.
- In a sexual population, Mendelian transmission means that only one of the two alleles in an individual, chosen at random at a locus, is transmitted to the offspring.



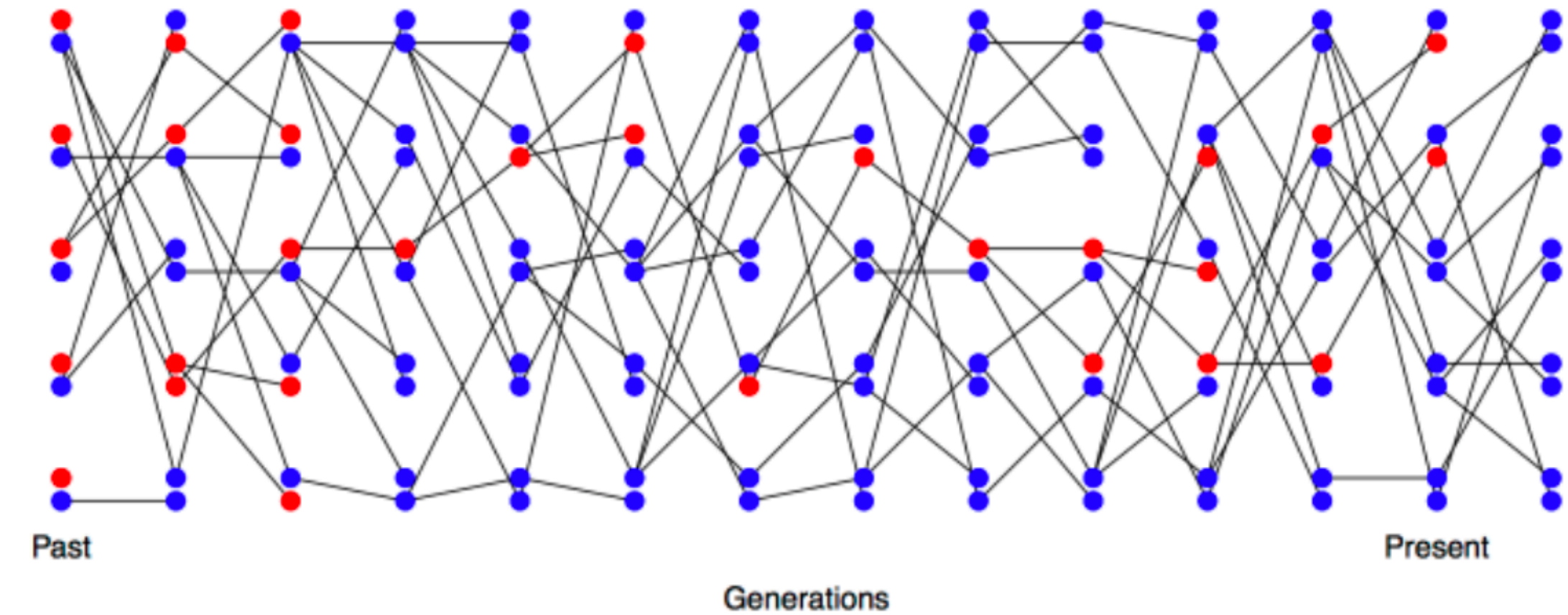
# Genetic drift

- If all alleles have equal chance of being inherited, the distribution of the number of alleles  $2N \times q'$  in the next generation is binomially distributed

$$P(q') = q^{2N \times q'} (1 - q)^{2N - 2N \times q'} \binom{2N}{2N \times q'}$$

$$\text{Var}(q') = \frac{1}{4N^2} \text{Var}(q' \times 2N) = \frac{1}{4N^2} 2N(1 - q)q$$

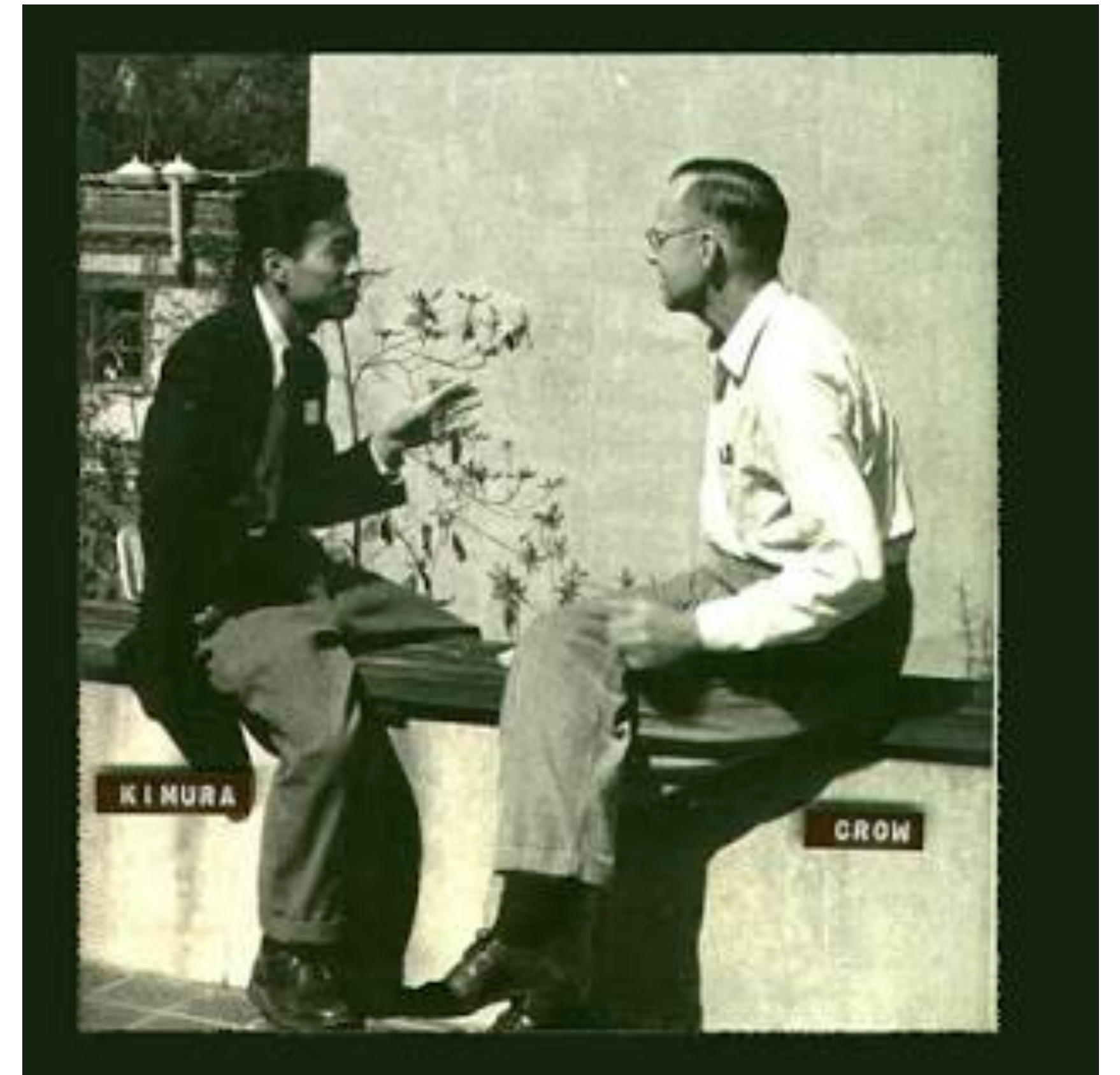
$$= \frac{1}{2N} (1 - q)q$$





# The neutral theory of molecular evolution

- Patterns of molecular polymorphism within species and substitution between species can be well understood by supposing that the vast majority of these molecular polymorphisms and substitutions were neutral alleles, whose dynamics were just subject to the vagaries of genetic drift and mutation.
- Beneficial mutations are rare enough that their contribution to the bulk of polymorphism or divergence can be largely ignored
- What mutations may be neutral?



# The neutral theory of molecular evolution

- Changes in non-coding DNA
- Synonymous changes in coding regions
- Non-synonymous changes that don't have a strong effect on the functional properties of the amino acid encoded
- An amino-acid change with phenotypic consequences, but little relevance to fitness

# Empirical measurement of the effect of mutations

- Distribution of fitness effects (DFE) mutations

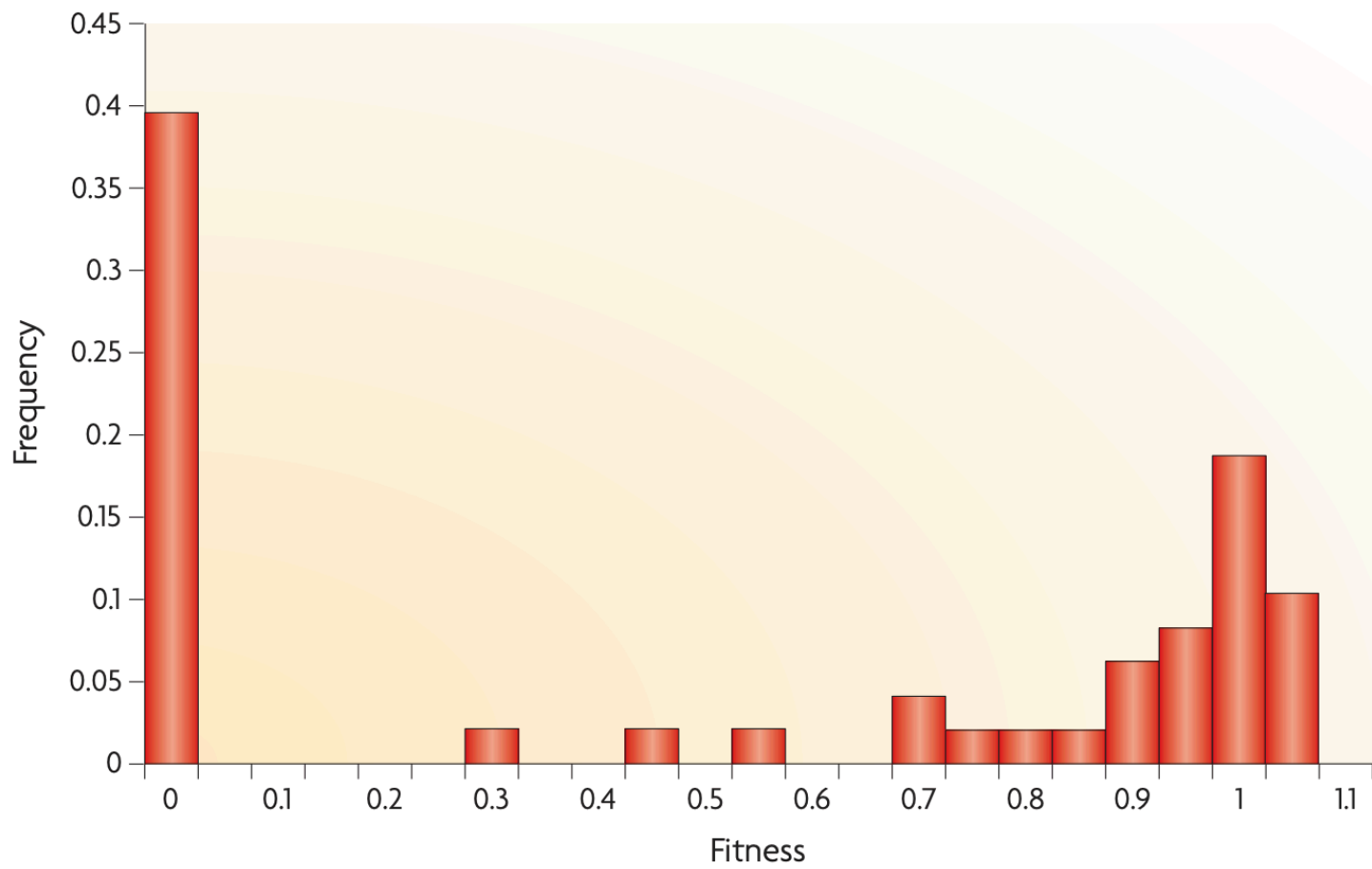


Figure 1 | **The distribution of fitness effects of random mutations in vesicular stomatitis virus.** In this experiment, random mutations were introduced into the virus, and the fitnesses of the mutants were compared against the unmutated wild type. A fitness of less than one indicates that the mutant was less fit than the wild type, so the mutation was deleterious. A fitness of zero indicates that no mutated progeny were recovered, and that the mutation was therefore lethal. Data from REF. 15.

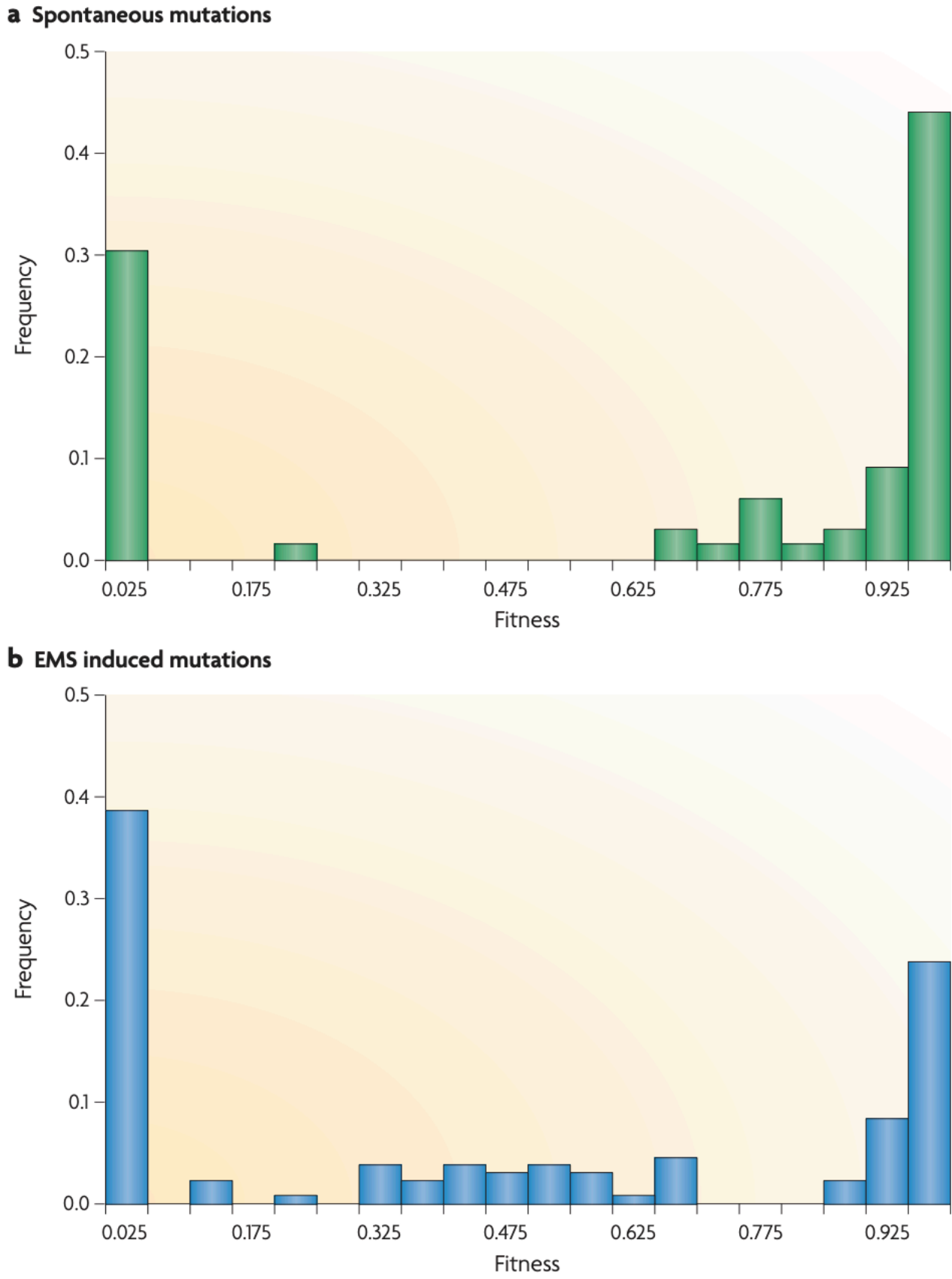
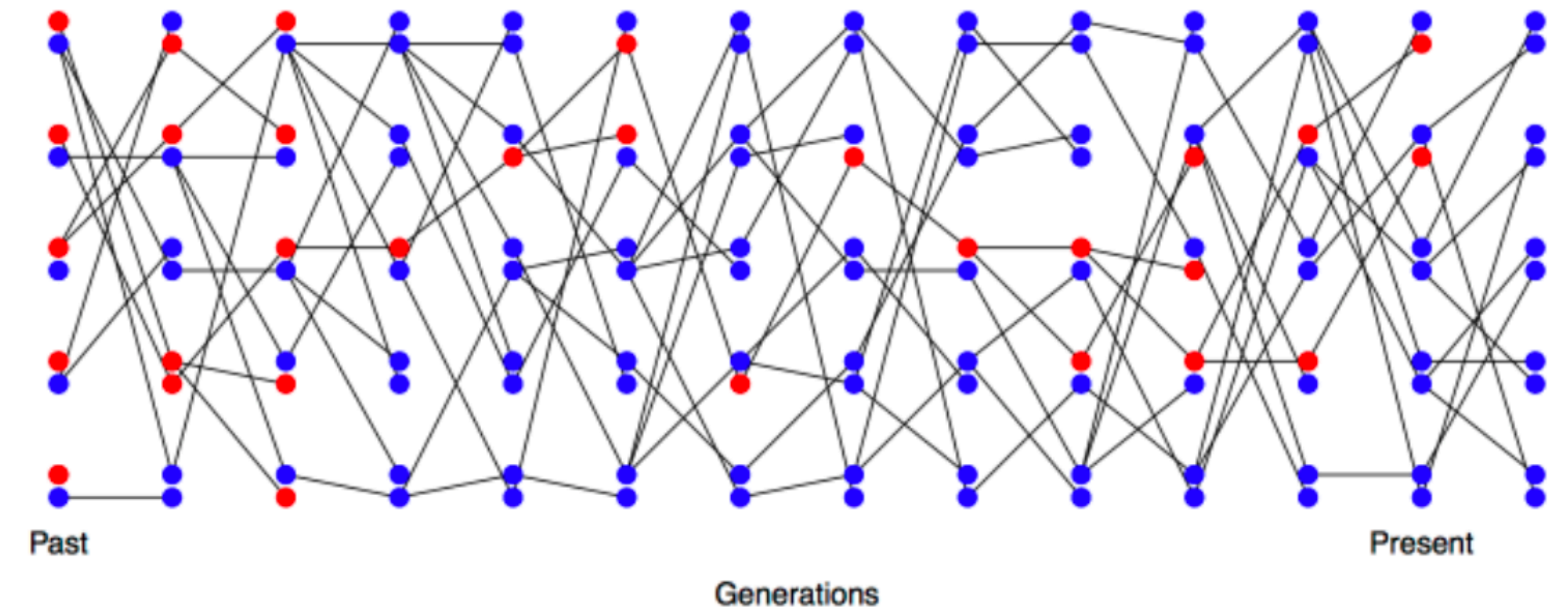


Figure 2 | **The distribution of fitnesses among yeast lines.** Diploid yeast lines were either allowed to accumulate spontaneous mutations (panel a) or were subject to chemical mutagenesis using ethylmethane sulphonate (EMS) (panel b). After a period of inbreeding, the cells were made to undergo meiosis and the growth of the meiotic products was measured. Data from REF. 17.



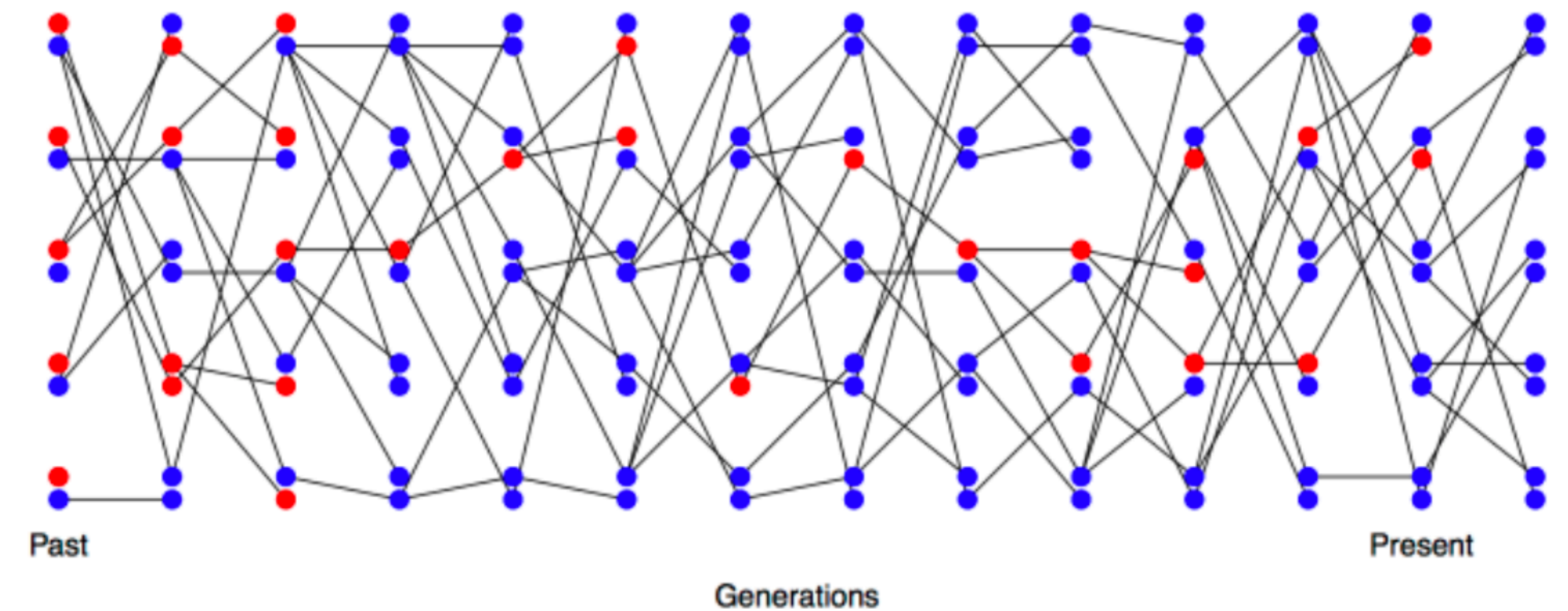
# Loss of heterozygosity due to drift

- Genetic drift will, in the absence of new mutations, slowly purge our population of neutral genetic diversity
- This leads to the loss of heterozygosity over time



# Loss of heterozygosity due to drift

- Evolution under **Wright-Fisher model** with no selection
  - Randomly mating population of a *constant size*  $N$  diploid individuals ( $2N$  alleles)
  - Discrete generations
  - Random mating
  - A locus segregating for two alleles that are neutral with respect to each other.
- No Mutation
- In generation  $t$  our current level of heterozygosity is  $H_t$
- **What is the level of heterozygosity in generation  $t + 1$ ?**





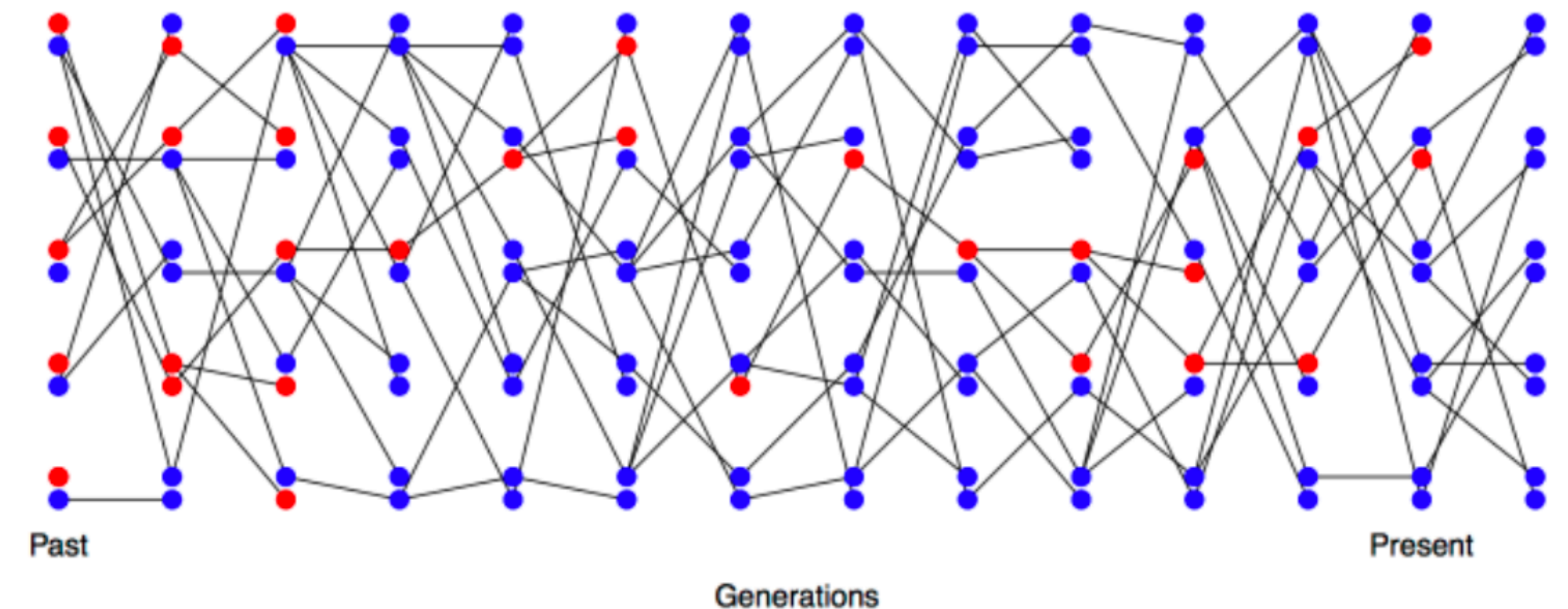
# Loss of heterozygosity due to drift

- The probability that our two alleles have the same parental allele in the proceeding generation is  $\frac{1}{2N}$
- The probability that they have different parental alleles is  $1 - \frac{1}{2N}$
- Apply the law of total probability, the *expected* heterozygosity in generation  $t + 1$  is:

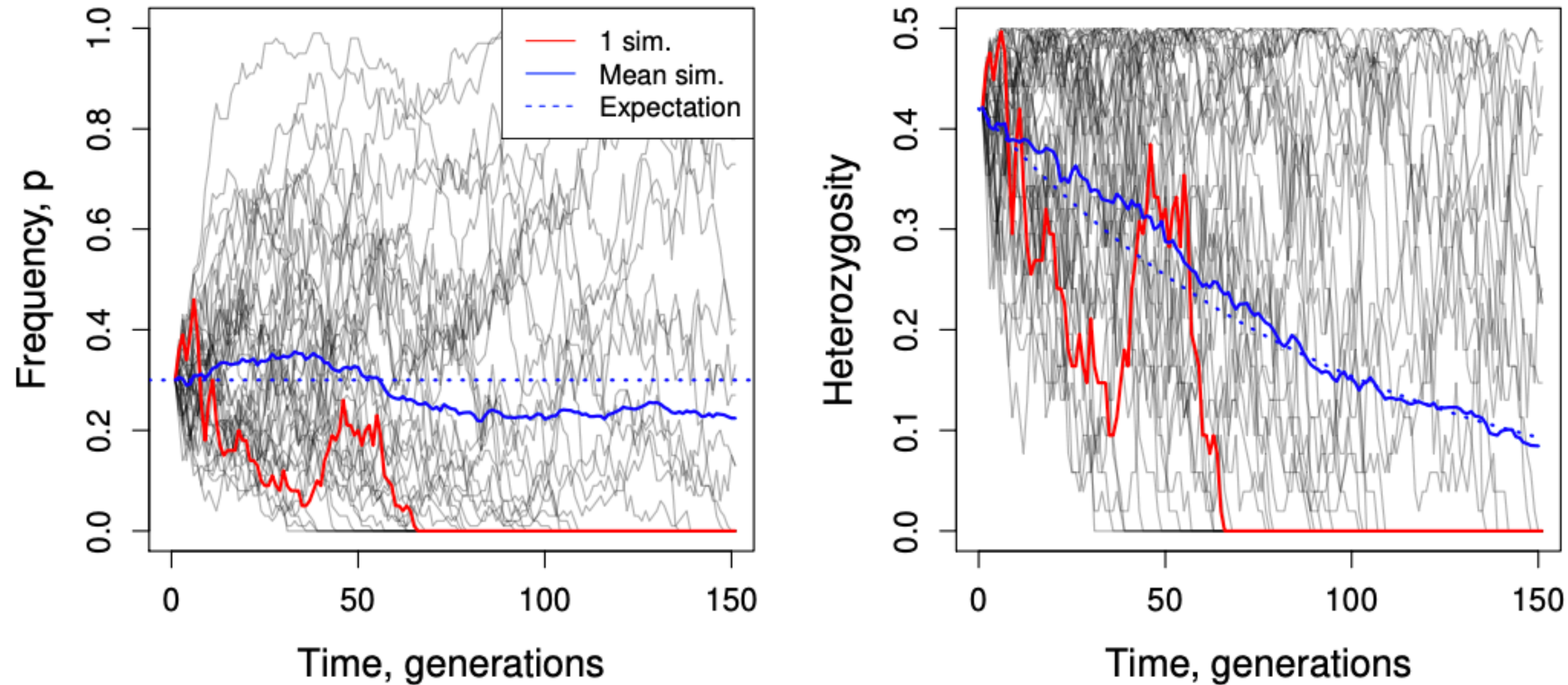
$$\bullet H_{t+1} = \frac{1}{2N} \times 0 + \left(1 - \frac{1}{2N}\right) \times H_t$$

- We can write down a recursive formula

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



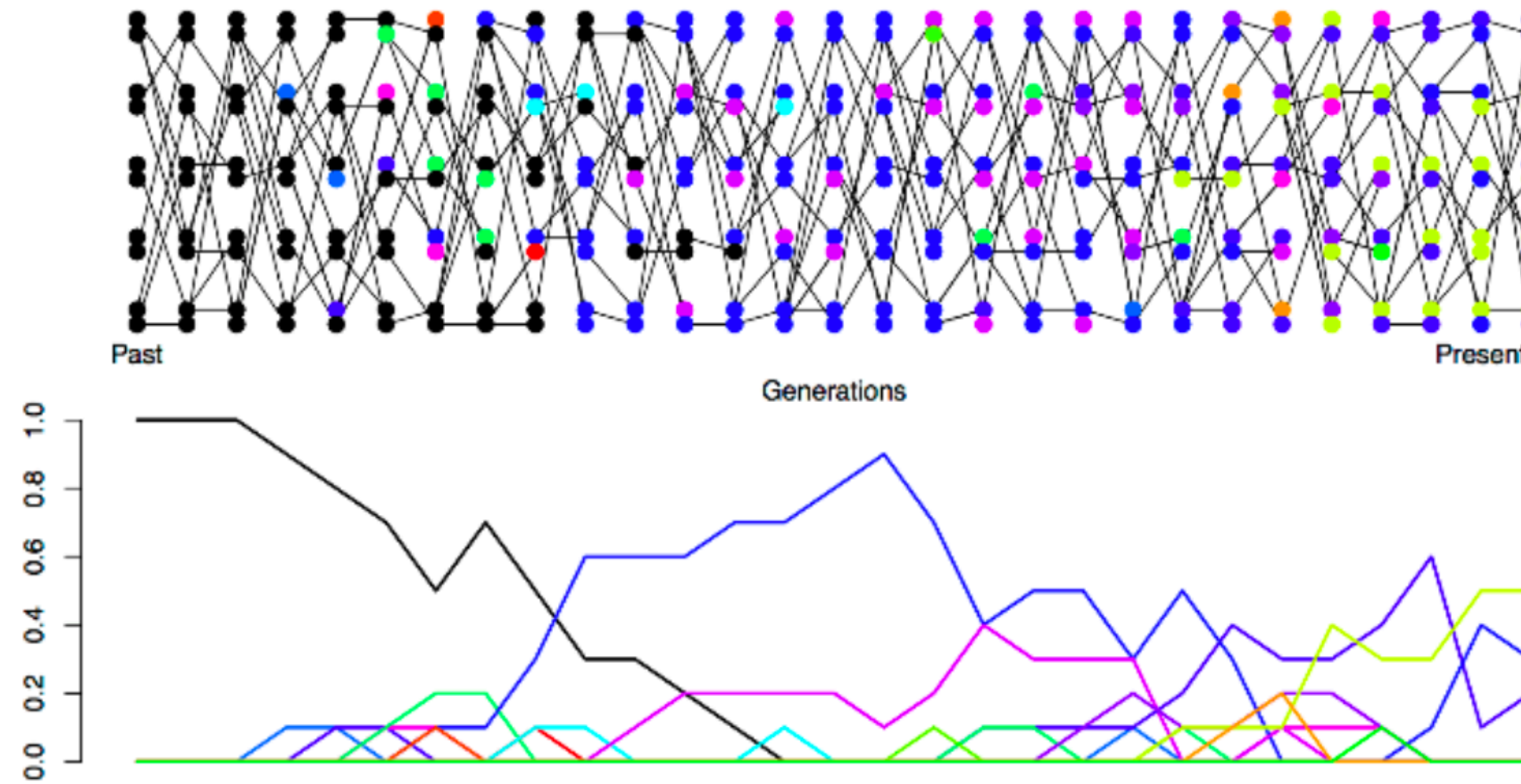
# Loss of heterozygosity due to drift



- Change in allele frequency and loss of heterozygosity over time for 40 replicates. Simulations of genetic drift in a diploid population of 50 individuals, in the absence of new mutations.
- Started with 40 independent, biallelic loci each with an initial allele at 30% frequency. The left panel shows the allele frequency over time and the right panel shows the heterozygosity over time, the mean (blue) matches our equation



# Levels of diversity maintained by the mutation-drift balance



- Mutation-drift balance. A diploid population of 5 individuals. In the first generation everyone has the same allele (black). Each generation the transmitted allele can mutate and we generate a new colour.
- Bottom plot shows frequency of alleles in our population over time. The mutation rate we use is very high, simply to maintain diversity in this small population.

# Mutation rate

- The per base pair mutation rate in humans is around  $1.5 \times 10^{-8}$  per generation
- A) Your autosomal genome is ~3 billion base pairs long. You have two copies, the one you received from your mum and one from your dad. What is the average (i.e. the expected) number of mutations that occurred in the transmission from your mum and your dad to you?
- B) The current human population size is ~7 billion individuals. How many times, at the level of the entire human population, is a single base-pair mutated in the transmission from one generation to the next?



# Levels of heterozygosity maintained by mutation-drift balance

- We'd like to derive the expected heterozygosity rate under random mating, and mutation-drift balance
- Expected heterozygosity ( $H$ ) = Prob two random alleles are different
- P of a random pair of alleles share a parental allele (have coalesced) in the preceding generation  $= \frac{1}{2N}$
- P of them fail to coalesce  $= 1 - \frac{1}{2N}$
- P a mutation occur in a generation  $= \mu$
- P no mutation occurring  $= 1 - \mu$
- Putting these together:
  - P of two randomly sampled alleles coalesce 2 generations in the past and are identical  
$$= \left(1 - \frac{1}{2N}\right) \frac{1}{2N} (1 - \mu)^4$$

# Levels of heterozygosity maintained mutation-drift balance

- More generally, the probability that our alleles coalesce in generation  $t + 1$  and are identical due to no mutation

$$\mathbb{P}(\text{coal. in } t+1 \text{ \& no mutations}) = \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^t (1 - \mu)^{2(t+1)}$$

- 

$$\mathbb{P}(\text{coal. in } t+1 \text{ \& no mutations}) \approx \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^t (1 - \mu)^{2t}$$

-

# Levels of heterozygosity maintained mutation-drift balance

- Assuming that  $1/(2N) \ll 1$  and  $\mu \ll 1$ . This allows us to approximate the geometric decay as an exponential decay

$$\begin{aligned}\mathbb{P}(\text{coal. in } t+1 \text{ \& no mutations}) &\approx \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^t (1 - \mu)^{2t} \\ &\approx \frac{1}{2N} e^{-t/(2N)} e^{-2\mu t} \\ &= \frac{1}{2N} e^{-t(2\mu + 1/(2N))}\end{aligned}$$

- Two alleles could coalesce in generation  $t = 1, 2, 3 \dots$
- Apply law of total probability

$$\begin{aligned}\mathbb{P}(\text{coal. in any generation \& no mutations}) &\approx \mathbb{P}(\text{coal. in } t = 1 \text{ \& no mutations}) + \\ &\quad \mathbb{P}(\text{coal. in } t = 2 \text{ \& no mutations}) + \dots \\ &= \sum_{t=1}^{\infty} \mathbb{P}(\text{coal. in } t \text{ generations \& no mutation})\end{aligned}$$

- Approximate sum with integral:

$$\frac{1}{2N} \int_0^{\infty} e^{-t(2\mu + 1/(2N))} dt = \frac{1/(2N)}{1/(2N) + 2\mu}$$

- equilibrium heterozygosity in a population at equilibrium between mutation and drift

$$H = \frac{2\mu}{1/(2N) + 2\mu} = \frac{4N\mu}{1 + 4N\mu}$$

$$\theta = 4N\mu$$



- What's the intuition of our eqn(4.13), well the probability that any event happens in a particular generation is  $P(\text{mutation or coalescence}) \approx 1/(2N) + 2\mu$ , so conditional 3 on an event happening the probability that it is a mutation is  $P(\text{mutation} \mid \text{mutation or coalescence}) = 2\mu / (1/(2N) + 2\mu)$ .
- So all else being equal, species with larger population sizes should have proportionally higher levels of neutral polymorphism

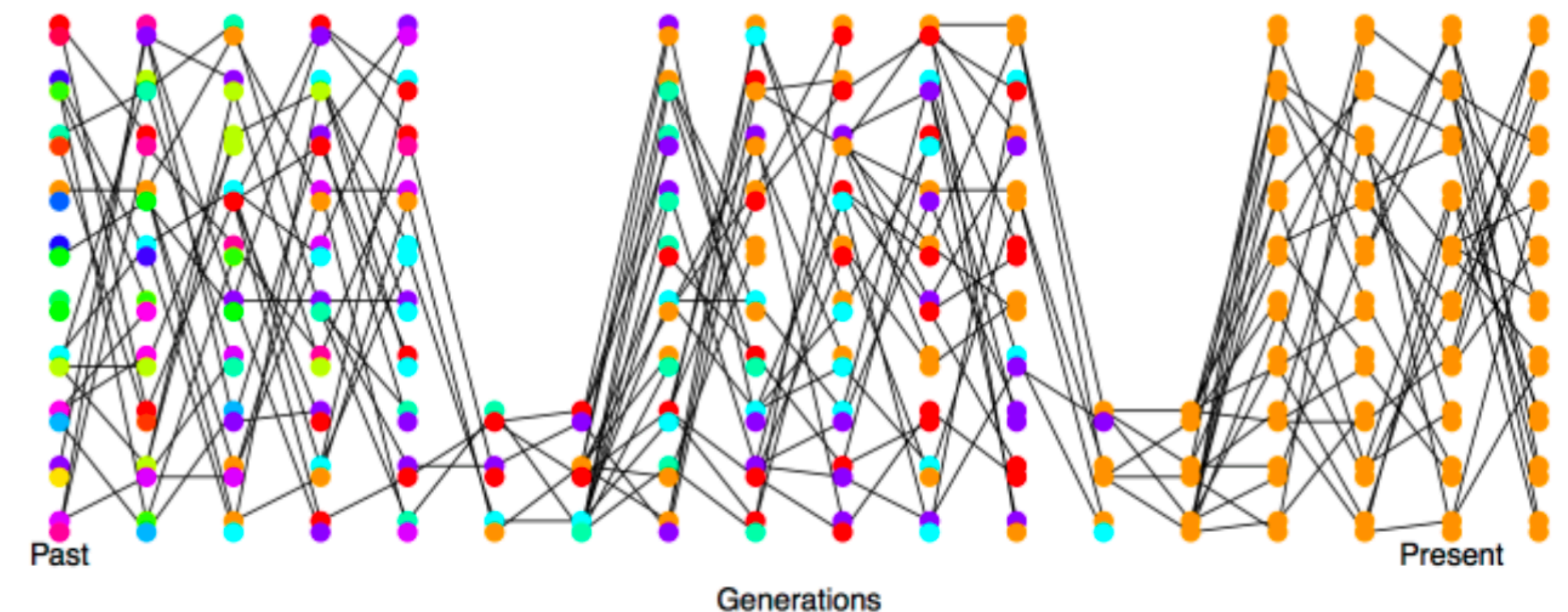
- 





# The effective population size

- populations rarely conform to our assumptions of being constant in size with low variance in reproductive success.
- Real populations experience dramatic fluctuations in size
- There is often high variance in reproductive success.
- The effective population size  $N_e$ :
  - Size of an idealized population that share common evolutionary dynamics as our real population



# Varying population size

- Consider a diploid population of variable size, whose size is  $N_t$  when traced  $t$  generations into the past.
- The probability our pairs of alleles have not coalesced by generation  $t$  is given by

- $$\prod_{i=1}^t \left(1 - \frac{1}{2N_i}\right)$$

- The rate of loss of heterozygosity is equivalent to a population of constant size

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

- The harmonic mean of the varying population size

# Variance in reproductive success

- Variance in reproductive success will also affect our effective population size. Even if our population has a large constant size  $N$  individuals, if only small proportion of them get to reproduce, then the rate of drift will reflect this much smaller number of reproducing individuals.
- $N_F$  of females get to reproduce and  $N_M$  males get reproduce.
- our probability of coalescence in the preceding generation is

- $$\frac{1}{4} \left( \frac{1}{2N_M} \right) + \frac{1}{4} \left( \frac{1}{2N_F} \right)$$

- $$N_e = \frac{4N_F N_M}{N_F + N_M}$$