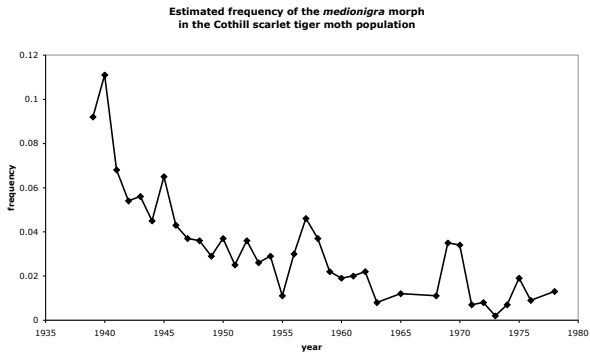


## Randomly fluctuating color polymorphism in the scarlet tiger moth



Figure 1. The scarlet tiger moth, *Panaxia dominula*. Top, the typical form (homozygote). Centre, *f. medionigra* (classical: CM) (heterozygote). Note the absence of the central yellow forewing spot in this specimen. Occasionally a small one is present. Bottom, *f. bimaculata* (homozygote). Photograph: Phil Hurst (NHM). Used with permission of Academic Press and editor, *The Linnean*.



sources: Fisher & Ford (1947), O'Hara (2005)

## The Wright-Fisher Model: Demographic Stochasticity and Genetic Drift (Fisher 1922, 1930; Wright 1931)

### Assumptions:

- 1 **Non-overlapping generations:** adults reproduce and then die.
- 2 **Constant population size:** exactly  $N$  haploid individuals survive to adulthood.
- 3 **No mutation or migration.**
- 4 **Neutrality:** survival and reproduction are independent of genotype.
- 5 **Random sampling:** each individual alive in generation  $t + 1$  independently chooses its parent uniformly at random and with replacement from the preceding generation.

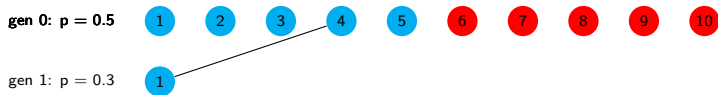
Equivalently: each individual alive in generation  $t$  gives birth to a very large number of offspring, each of which is equally likely to survive to reproductive maturity in generation  $t + 1$ .

## Simulation of the Wright-Fisher Model ( $N = 10$ )

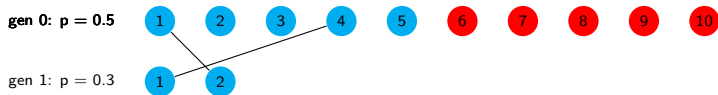
gen 0:  $p = 0.5$



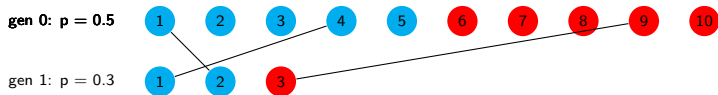
## Simulation of the Wright-Fisher Model ( $N = 10$ )



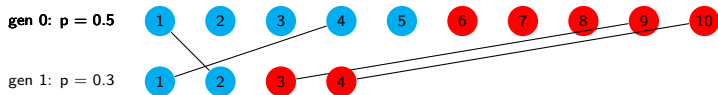
## Simulation of the Wright-Fisher Model ( $N = 10$ )



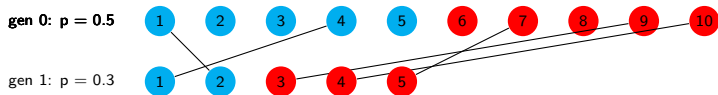
## Simulation of the Wright-Fisher Model ( $N = 10$ )



## Simulation of the Wright-Fisher Model ( $N = 10$ )

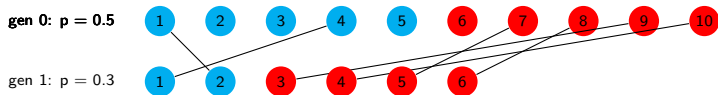


## Simulation of the Wright-Fisher Model ( $N = 10$ )

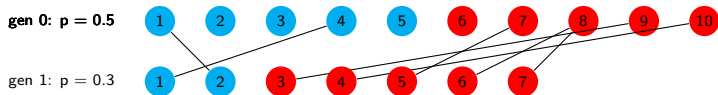




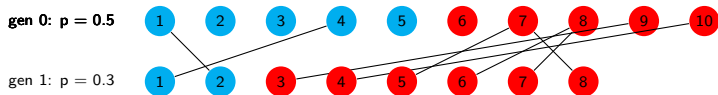
## Simulation of the Wright-Fisher Model ( $N = 10$ )



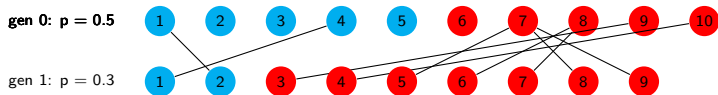
## Simulation of the Wright-Fisher Model ( $N = 10$ )



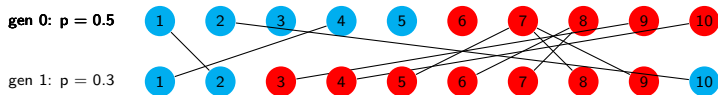
## Simulation of the Wright-Fisher Model ( $N = 10$ )



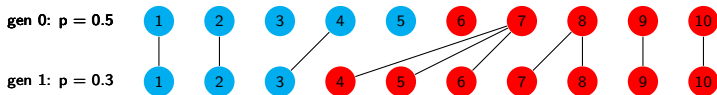
## Simulation of the Wright-Fisher Model ( $N = 10$ )



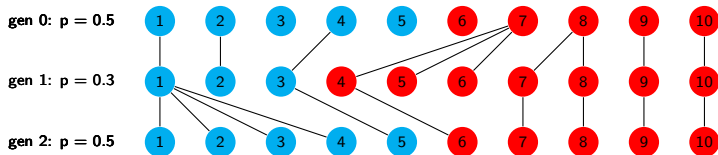
## Simulation of the Wright-Fisher Model ( $N = 10$ )



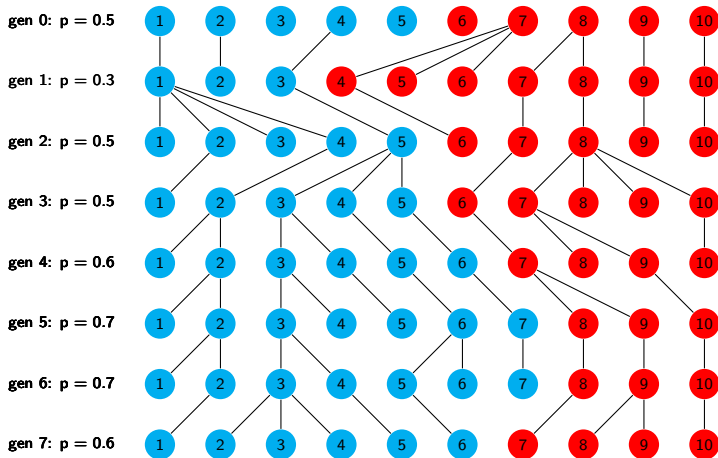
## Simulation of the Wright-Fisher Model ( $N = 10$ )



## Simulation of the Wright-Fisher Model ( $N = 10$ )

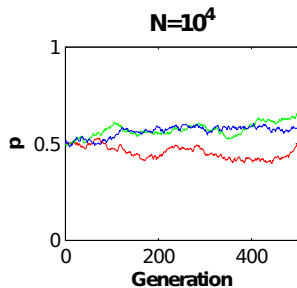
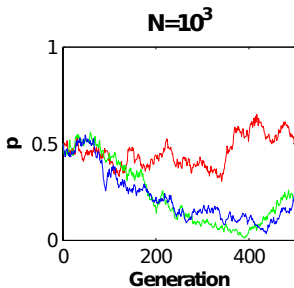
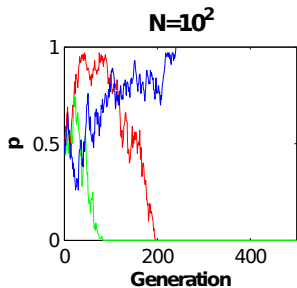


## Simulation of the Wright-Fisher Model (N = 10)





**Genetic drift: Fluctuations in the genetic composition of the population due to random variation in survival and reproduction.**



Many factors may contribute to genetic drift, including:

- random variation in survival
- random variation in number of mates
- random variation in fecundity
- environmental variation affecting the above
- selection at unlinked sites
- random segregation (Mendel's first law)

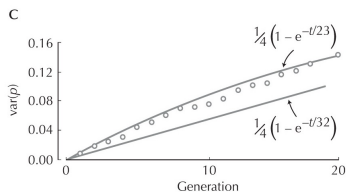
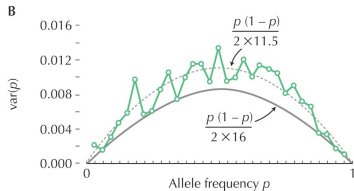
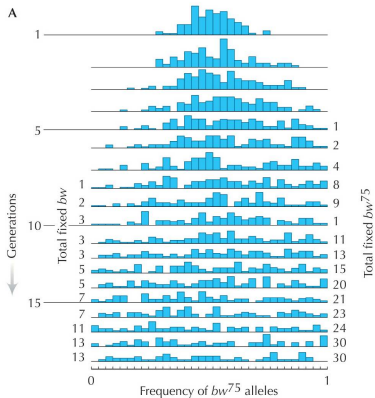
## Statistical Properties of genetic drift under the Wright-Fisher model

$$\begin{aligned}\Delta p_t &= p_{t+1} - p_t \\ \mathbb{E}[\Delta p_t] &= 0 \\ \text{Var}(\Delta p_t) &= \frac{p_t(1-p_t)}{N}.\end{aligned}$$

Thus, some key predictions of the **neutral** Wright-Fisher model are:

- The expected allele frequency is constant.
- The variance of the allele frequency fluctuations is inversely proportional to the population size.
- Genetic drift is **weaker** and **slower** in larger populations.

## Example: Genetic drift at an eye color locus in experimental populations of *D. melanogaster*



Source: P. Buri (1956) "Gene frequency in small populations of mutant *Drosophila*". *Evolution* 10: 367-402.

## Loss of Variation in Finite Populations

Although on average genetic drift does not change allele frequencies, in the long run it will reduce the amount of variation in a population. Let  $\mathcal{G}_t$  be the probability that two randomly sampled chromosomes are **identical-in-state** and let  $\mathcal{H}_t = 1 - \mathcal{G}_t$ .

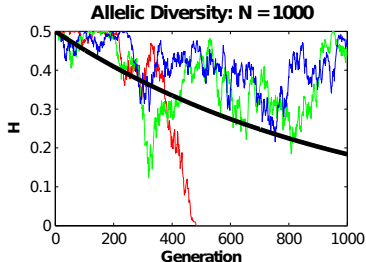
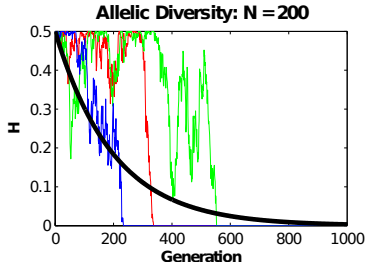
### Loss of genetic variation under the Wright-Fisher model

$$\begin{aligned}\mathcal{G}_t &= \frac{1}{N} + \left(1 - \frac{1}{N}\right) \mathcal{G}_{t-1} \\ \mathcal{H}_t &= \left(1 - \frac{1}{N}\right)^t \mathcal{H}_0.\end{aligned}$$

**Note:** These results hold for a haploid Wright-Fisher model. If the population contains  $N$  diploid individuals, then we must replace  $N$  in the above equations by  $2N$ .

This formula has several important consequences:

- In the long term, the expected value of  $\mathcal{H}_t$  will decrease to 0 as  $t \rightarrow \infty$ .
- On average, variation is lost more rapidly in small populations than in large populations.
- In particular, it takes approximately  $N \cdot \ln(2)$  generations for heterozygosity to be halved.



## Fixation of Neutral Alleles

In the absence of mutation and immigration, eventually all but one of the alleles initially present in a population will be lost through genetic drift. In this case, we say that the sole surviving allele has been **fixed** in the population. The W-F model can be used to explore some of the statistical properties of this process.

- The probability that an allele is eventually fixed in the population is **equal to its initial frequency**:

$$\mathbb{P}(A \text{ is eventually fixed}) = p.$$

- The mean time until one of the alleles is fixed is proportional to the population size. With two alleles with initial frequencies  $p$  and  $1 - p$ , respectively, this is

$$\mathbb{E}[T_{fix}] \approx -2N\{p \ln(p) + (1 - p) \ln(1 - p)\}.$$

## Neutral Substitution Processes

Suppose that new mutations appear at rate  $\mu$  per generation per individual in a haploid population of size  $N$  and assume that each mutation is neutral.

- The expected number of new mutations produced in each generation is  $N\mu$ .
- Provided that each new mutation initially appears in a single individual, its initial frequency is  $1/N$ .
- In this case, the probability that a new mutation will eventually be fixed in the population is also  $1/N$ .

If  $\mu \ll 1/N$ , then typically each new mutation will either be lost or go to fixation before the next mutation appears in the population. It follows that the rate at which new mutations are fixed in the population is

$$\text{substitution rate} = N\mu \times \frac{1}{N} = \mu.$$



This calculation has several important consequences:

- The substitution rate of neutral alleles is equal to the mutation rate.
- Divergence at neutral loci (in the sense of fixation of neutral alleles) does not depend on population size.
- On the other hand, polymorphism (variation within populations) does depend on the population size, since this determines the strength of genetic drift in the W-F model.
- Differentiation of recently isolated populations also depends on population size, since this determines the rate at which the allele frequencies fluctuate away from one another.

## Simulation of the Wright-Fisher Model

Several different approaches are available for simulating the W-F model, depending on one's goals.

- **Direct method:** In this approach, one randomly chooses the parent of each individual alive in the present generation from the  $N$  individuals alive in the previous generation. This gives information about allele frequencies and the relationships between individuals in the population.
- **Binomial method:** In this case, the number of  $A$  alleles present in generation  $t + 1$  is determined by sampling a **binomial random variable** with parameters  $N$  and  $p = X_t/N$ :

$$\begin{aligned} X_{t+1} &\sim \text{Binomial}(N, p) \\ \mathbb{P}(X_{t+1} = k) &= \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}. \end{aligned}$$

## Fast Simulation of the Wright-Fisher Model

When  $N$  is large, the binomial method can be accurately approximated by a much faster algorithm that exploits two approximations for the binomial distribution.

- When  $p$  and  $1 - p$  are not too close to 0, say  $\min Np, N(1 - p) > 25$ , the binomial distribution can be approximated by a normal distribution:

$$X_t \sim \text{Normal}(Np, Np(1 - p)).$$

- If  $Np \leq 25$ , then the binomial distribution can be approximated by a Poisson distribution:

$$X_t \sim \text{Poisson}(Np).$$

- If  $N(1 - p) \leq 25$ , then the binomial distribution can be approximated by a Poisson distribution:

$$N - X_t \sim \text{Poisson}(N(1 - p)).$$

## Demography of the Wright-Fisher Model

Let  $\eta_i, i = 1, \dots, N$  be the random number of offspring contributed by the  $i$ 'th individual to the next generation. Under the Wright-Fisher model, these numbers are random variables with the following properties:

- Each of the offspring numbers is a binomial random variable:  
 $\eta_i \sim \text{Binomial}(N, 1/N)$ .
- When  $N$  is large, these are approximately Poisson-distributed:  $\eta_i \sim \text{Poisson}(1)$ :

$$\mathbb{P}(\eta_i = k) = \frac{1}{e} \frac{1}{k!}$$

- In particular, since  $\text{Var}(\eta_i) \approx 1$ , no individual is likely to contribute a large number of offspring to the next generation.
- However, since  $\mathbb{P}(\eta_i = 0) \approx 0.37$ , approximately 37% of the population will contribute no offspring to the next generation.

**Moral:** The Wright-Fisher model makes some fairly strong assumptions about demography and life history.

## The Moran Model: An Alternative Model of Genetic Drift (Moran 1958)

### Assumptions:

- 1 **Overlapping generations:** Individuals give birth and die at different times.
- 2 **Constant population size:** exactly  $N$  haploid adults are live at any given time.
- 3 **No mutation or migration.**
- 4 **Neutrality:** survival and reproduction are independent of genotype.
- 5 **Random sampling:** At each time step, an individual is randomly sampled to reproduce. They give birth to one offspring, which then replaces one of the  $N$  adults chosen uniformly at random.

**Remark:** Birth-death events can either occur at discrete or continuous times.