Simulating evolution using for loops

Brian Kissmer

USU Department of Biology

Sept. 10th, 2024

Today’s outline

1. The Wright-Fisher model
2. For loops

The Wright-Fisher model of evolution - History

Wright observed that allele frequency changes can occur (especially in small populations) via random removal of individuals or alleles from the population

While he coined this process the “Sewall Wright effect” (lol), the term came to be known as “genetic drift”

The Wright-Fisher model of evolution - History

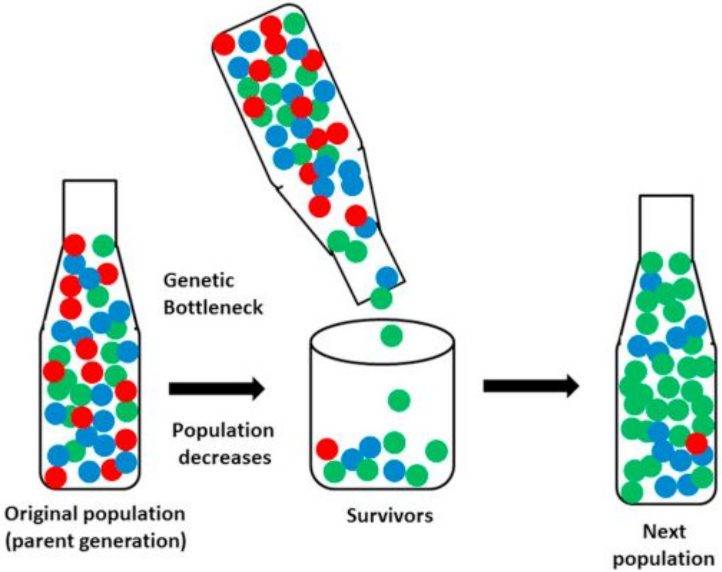
Ronald Fisher believed natural selection was the main force driving allele frequency change.

These two ideas were combined to establish the Wright-Fisher model of evolution, in which both drift and selection influence allele frequency change

Evolution via genetic drift

What exactly is genetic drift?

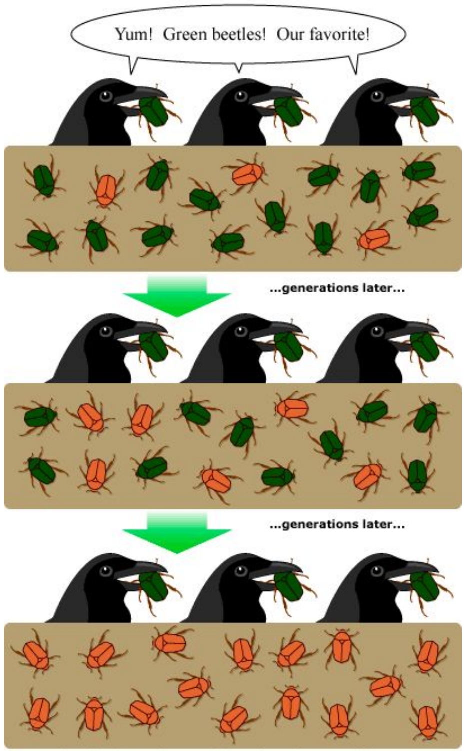
1. Variation in survival and reproduction that is <u>unrelated</u> to phenotypic or genetic differences among individuals
2. Akin to sampling error in a statistical context



Evolution via natural selection

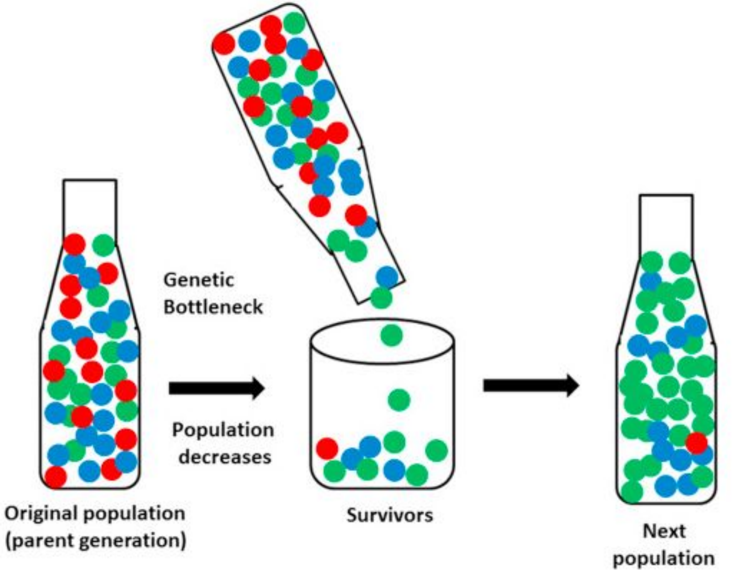
1. Individuals in a population vary in some trait
2. Those traits vary in their fitness in some environment
3. Variation for that trait is heritable and can be passed to offspring
4. Variation with higher fitness increases in frequency over time

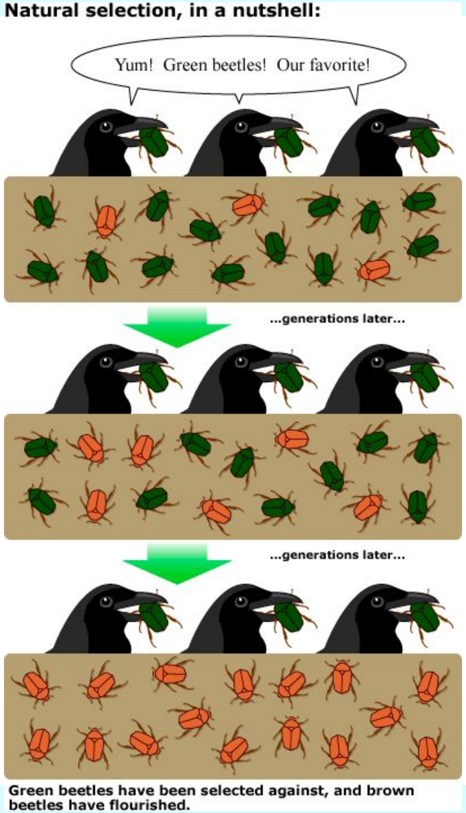
Natural selection, in a nutshell:



Green beetles have been selected against, and brown beetles have flourished.

Evolution via the Wright-Fisher model



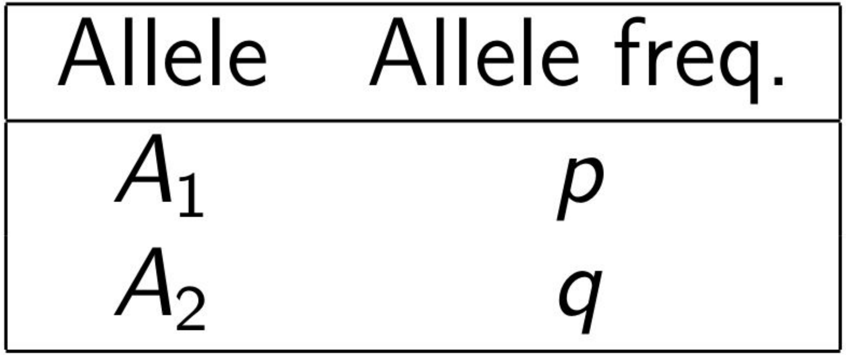


How can we model this process?

Modeling different scenarios gives us an idea about how changes in the importance of different parameters (population size, selection strength, etc.) can influence evolutionary outcomes

Allele frequency notation

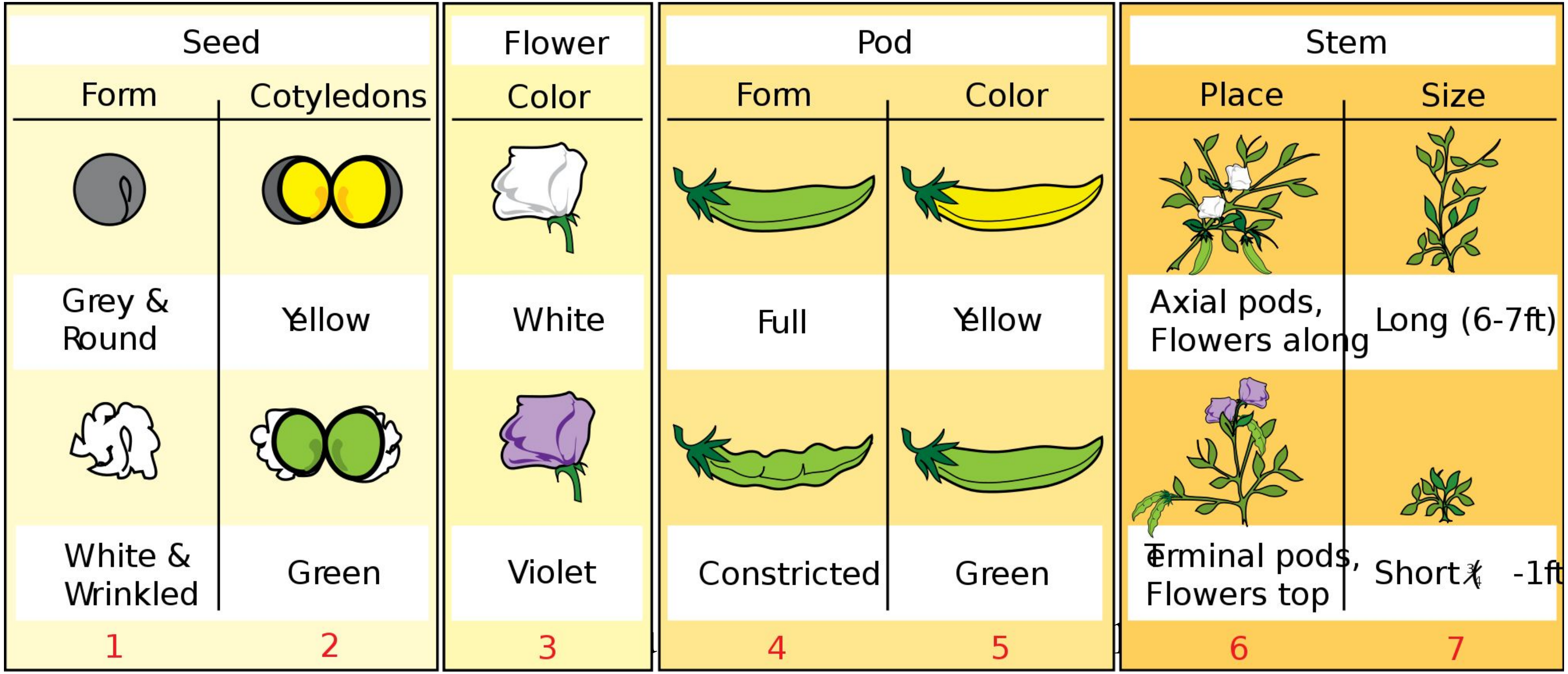
For a biallelic locus:



Allele	Allele freq.
A_1	p
A_2	q

The two allele frequencies (p and q) sum to 1, or $p + q = 1$

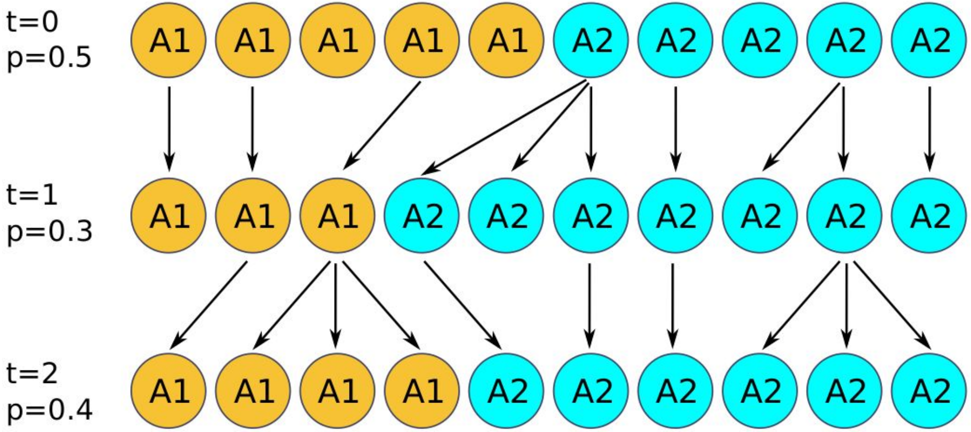
Allele frequency notation



How can we model this process? - Drift

Allele frequency change due to drift can be modeled as a binomial processes:

$$p_{t+1} \sim \text{binomial}(p_t, 2N)/2N$$



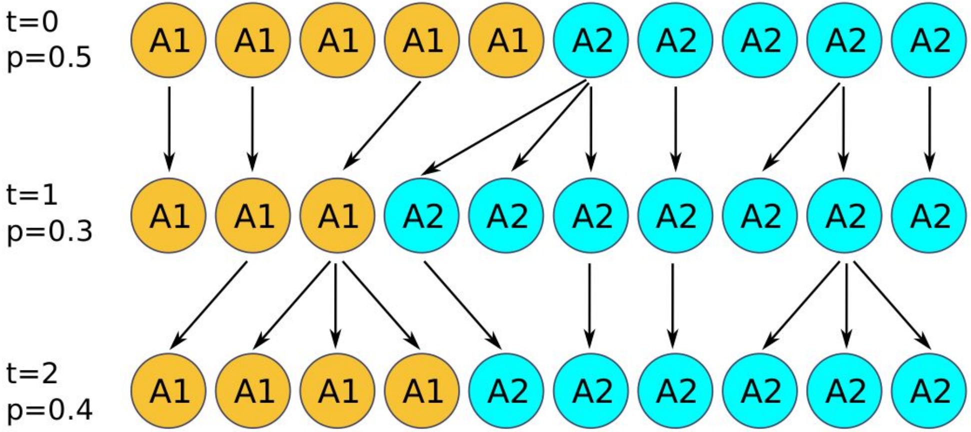
N is the number of individuals, 2N is the number of gene copies in a diploid population

How can we model this process? - Drift

Allele frequency change due to drift can be modeled as a binomial processes:

$$p_{t+1} \sim \text{binomial}(p_t, 2N)/2N$$

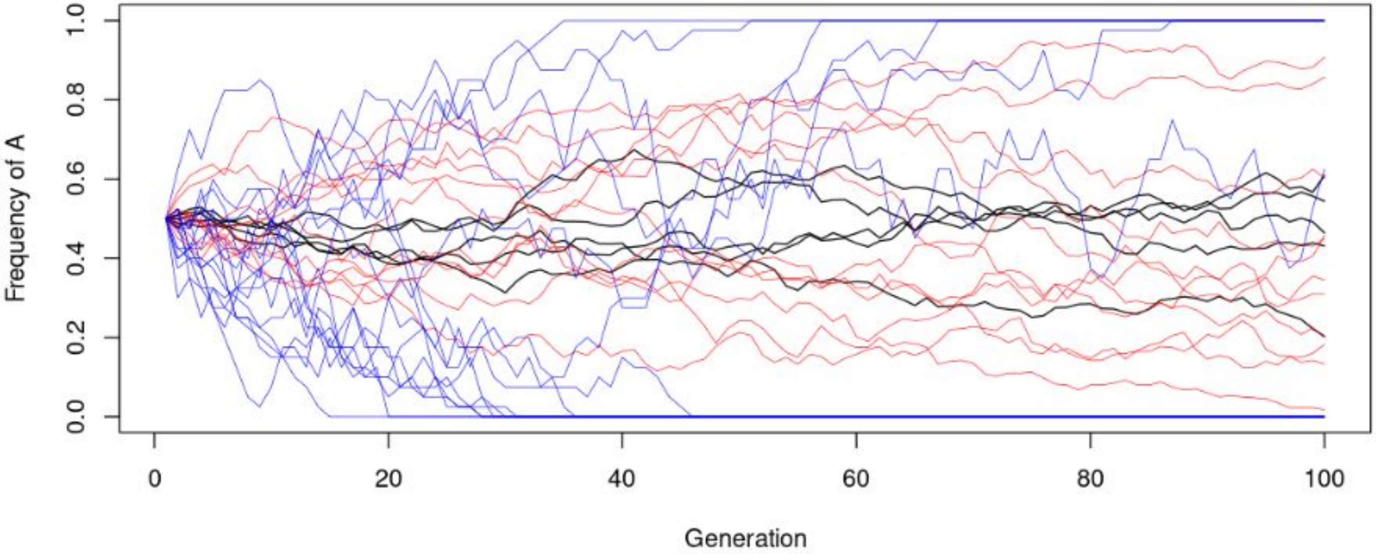
Break down the above equation with your group. What is the purpose of p_t and $2n$ within the parentheses. Can you figure out why we divide by 2N?



N is the number of individuals, 2N is the number of gene copies in a diploid population

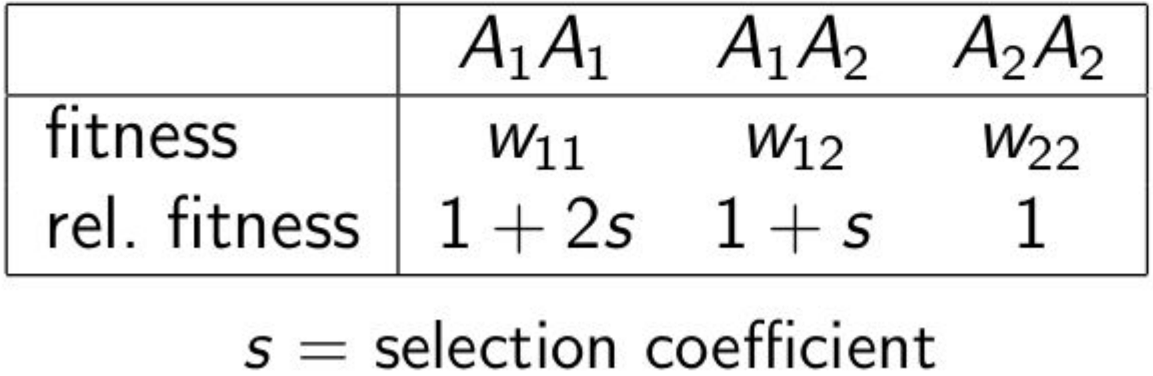
How can we model this process? - Drift

The effects of drift are influenced by population size. 2N = 20 (blue), 200 (red), 500 (black)



How can we model this process? - Selection

We can estimate the effect of selection from one generation to the next based on the relative fitness of our three diploid genotypes



	A_1A_1	A_1A_2	A_2A_2
fitness	w_{11}	w_{12}	w_{22}
rel. fitness	$1+2s$	$1+s$	1

$s =$ selection coefficient

How can we model this process? - Selection

Using relative fitness, we arrive at this approximate equation for evolution by natural selection:

$$\Delta_s p = sp(1 - p)$$

$$p_{t+1} = \Delta_s p_t + p_t$$

How can we model this process? - Selection and drift

To incorporate drift and natural selection into a single model, we can pull a binomial sample from our selection-adjusted allele frequency change:

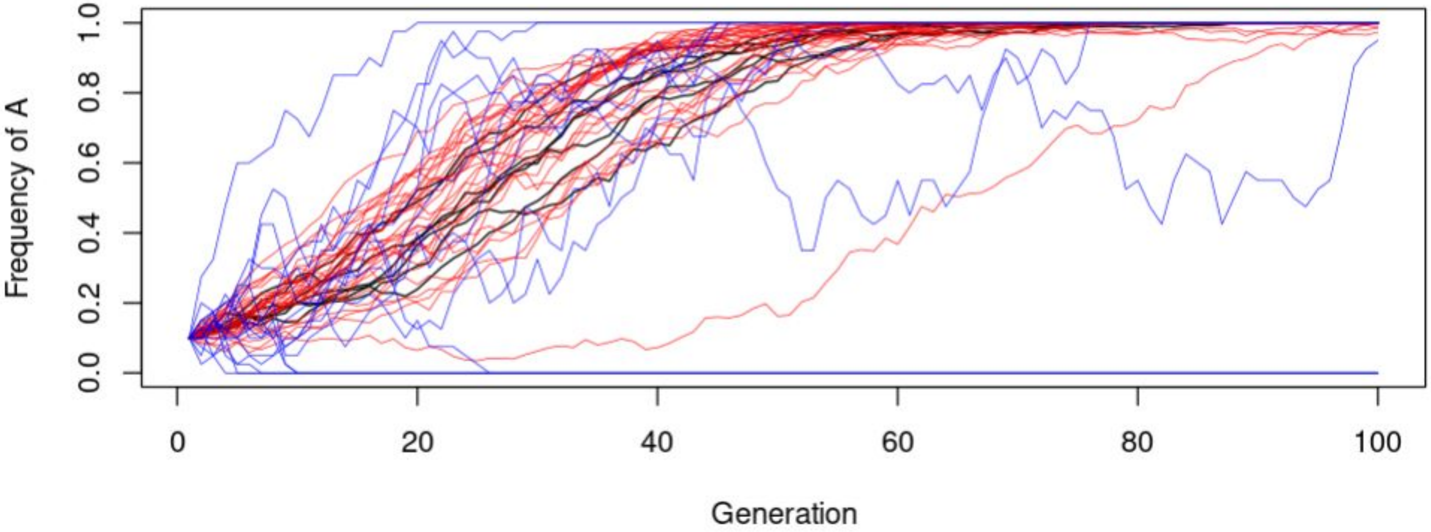
$$\Delta_s p_t = s p_t(1 - p_t)$$

$$p_{selection} = \Delta_s p_t + p_t$$

$$p_{t+1} = binomial(p_{selection}, 2N)/2N$$

How can we model this process? - Selection and drift

s = .1, 2N = 20 (blue), 200 (red), 500 (black)



Selection affects the mean allele frequency change, while drift affects the variance in allele frequency change

For loops

1. Now that we have a conceptual understanding of the Wight-Fisher model of evolution, let’s spend a bit of time on the R methods we will use to actually conduct our simulations. We’re going to be using most of the same methods you’ve already learned , but now we’ll be using ‘for loops’ to iterate over some variable (in this case, time).

Head to the “For Loops” handout