



2

Genetic drift

Mechanisms of evolution

There are four mechanisms able to cause allele frequency changes over time in a population

- Natural selection
- Genetic drift
- Mutation
- Migration

Hardy-Weinberg equilibrium

ASSUMPTIONS

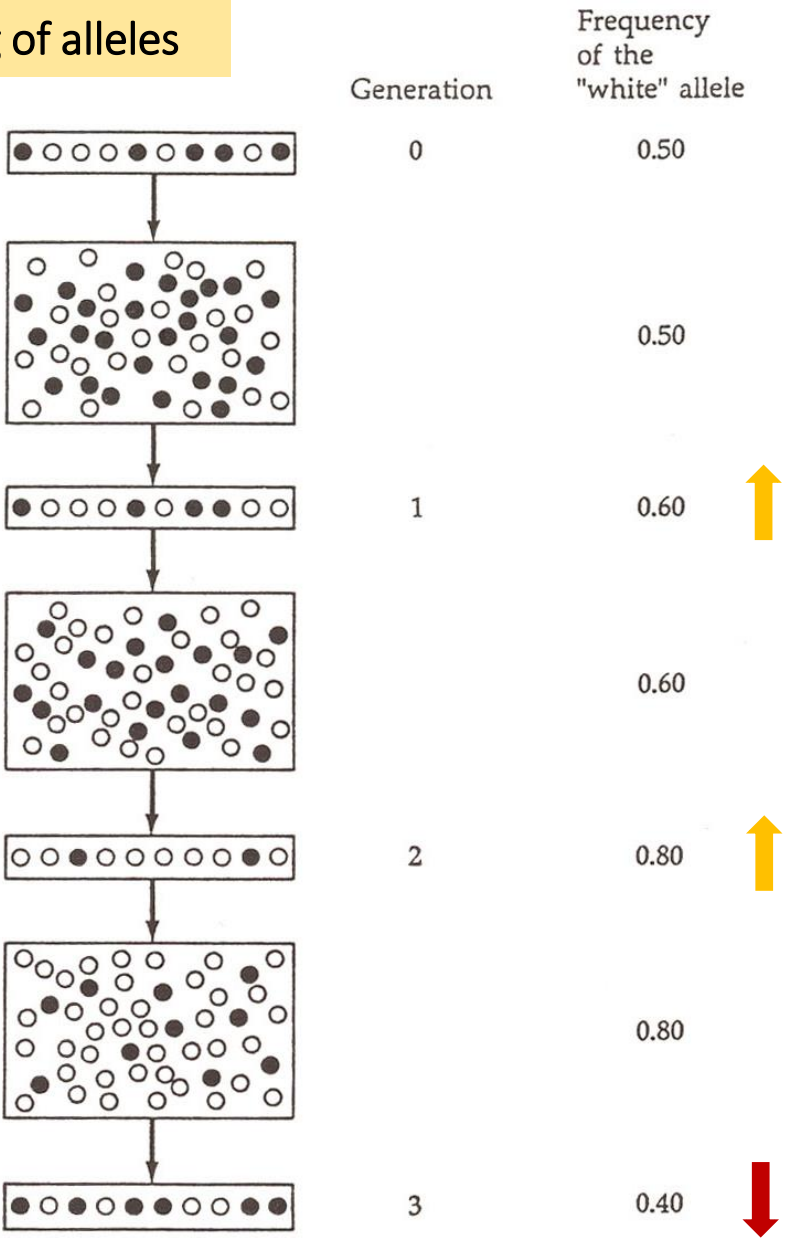
- Diploid organism
- Sexual reproduction
- Non-overlapping generations
- Random mating
- Equal allele frequencies in both sexes
- Large population size
- No migration
- No mutation
- No selection

Populations of finite size

Random sampling of alleles

The allele frequencies in the gamete pool are exactly the same as in the gamete-producing adults.

A random sample of 10 gametes is taken from the gamete pool.



Population size **N = 5**
Number of alleles **2N = 10**
Allelic frequencies **p = q = 0.5**

No drift

Finite population size

Allele frequencies = 0.5 in all successive generations

Allele frequencies fluctuate up and down

Genetic drift is a stochastic process

We can only predict the probability of each possible outcome in the next generation

Binomial distribution

The binomial distribution is used when:

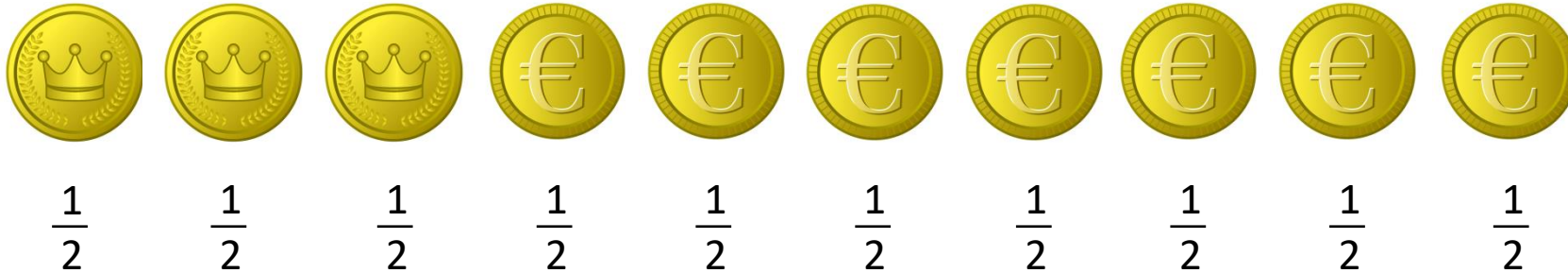
- there are two possible outcomes of a trial
- the probability of each outcome remains the same across all trials
- all trials are independent of each other

The probability of getting exactly k successes with p probability in n trials is:

$$P = \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k}$$

Binomial distribution

Which is the probability of getting 3 heads and 7 tails if you flip a coin 10 times?



Number of combinations
with 3 heads and 7 tails

Probability of getting tails

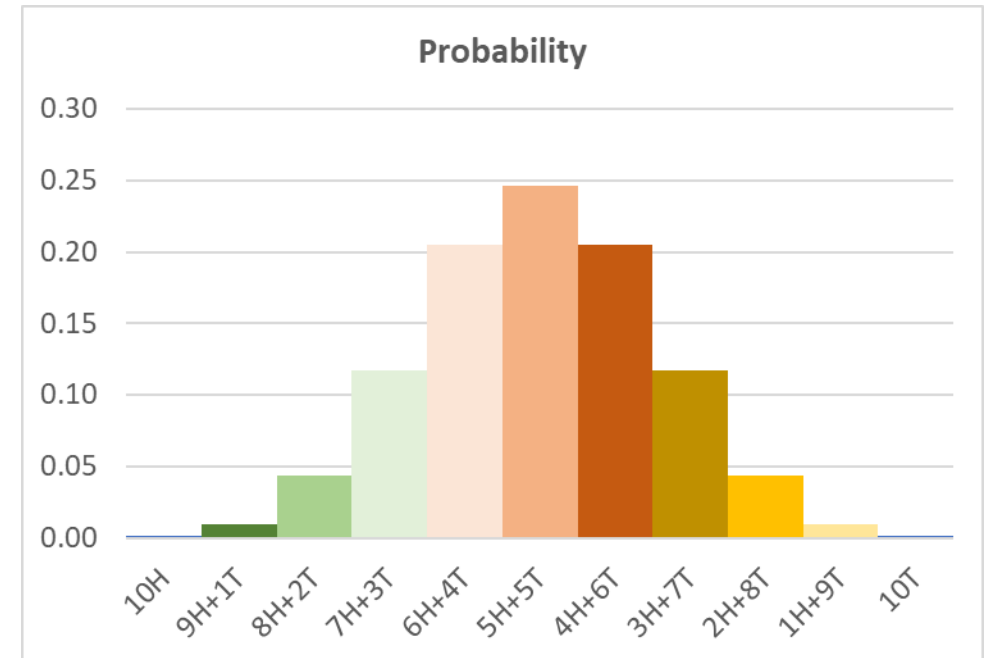
$$120 \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^7 = 0.1172$$

Probability of getting heads

$$\text{Number of combinations} = \frac{10!}{3!(10-3)!} = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2} = \frac{720}{6} = 120$$

Binomial distribution

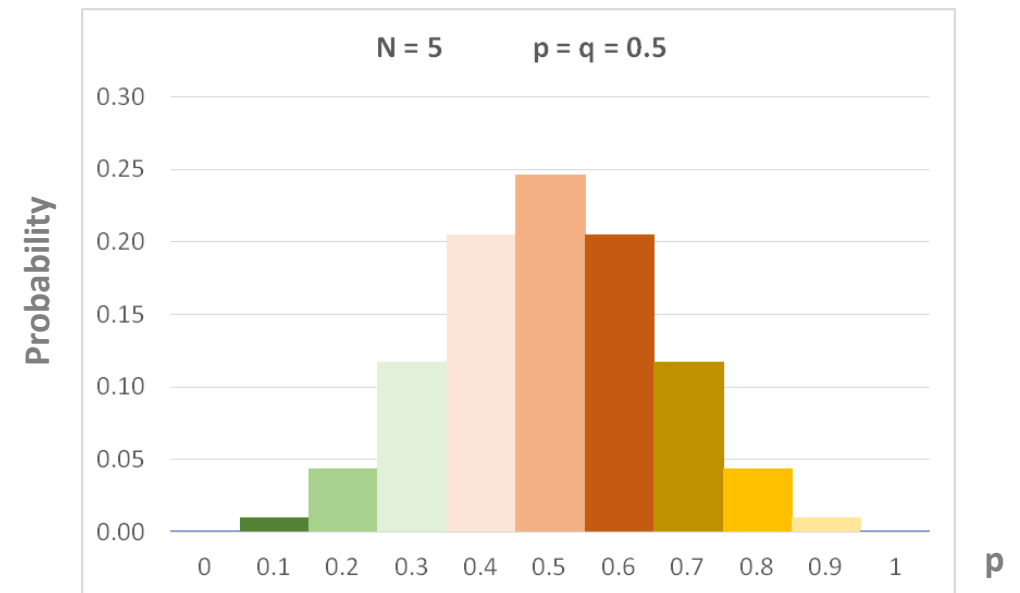
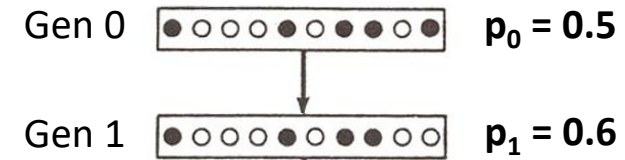
Possible results	Combinations	Probability
10 heads	1	0.0009765
9 heads + 1 tail	10	0.009765
8 heads + 2 tails	45	0.043945
7 heads + 3 tails	120	0.1172
6 heads + 4 tails	210	0.2051
5 heads + 5 tails	252	0.2461
4 heads + 6 tails	210	0.2051
3 heads + 7 tails	120	0.1172
2 heads + 8 tails	45	0.043945
1 head + 9 tails	10	0.009765
10 tails	1	0.0009765



~25% of times most likely result
~75% of times different result

Binomial distribution

Possible results	Combinations	Probability
10 white	1	0.0009765
9 white + 1 black	10	0.009765
8 white + 2 black	45	0.043945
7 white + 3 black	120	0.1172
6 white + 4 black	210	0.2051
5 white + 5 black	252	0.2461
4 white + 6 black	210	0.2051
3 white + 7 black	120	0.1172
2 white + 8 black	45	0.043945
1 white + 9 black	10	0.009765
10 black	1	0.0009765



$$P = \frac{2N!}{k! (2N-k)!} p^k q^{2N-k}$$

Binomial – Example to calculate

We have a population of 10 individuals with two alleles in a gene.

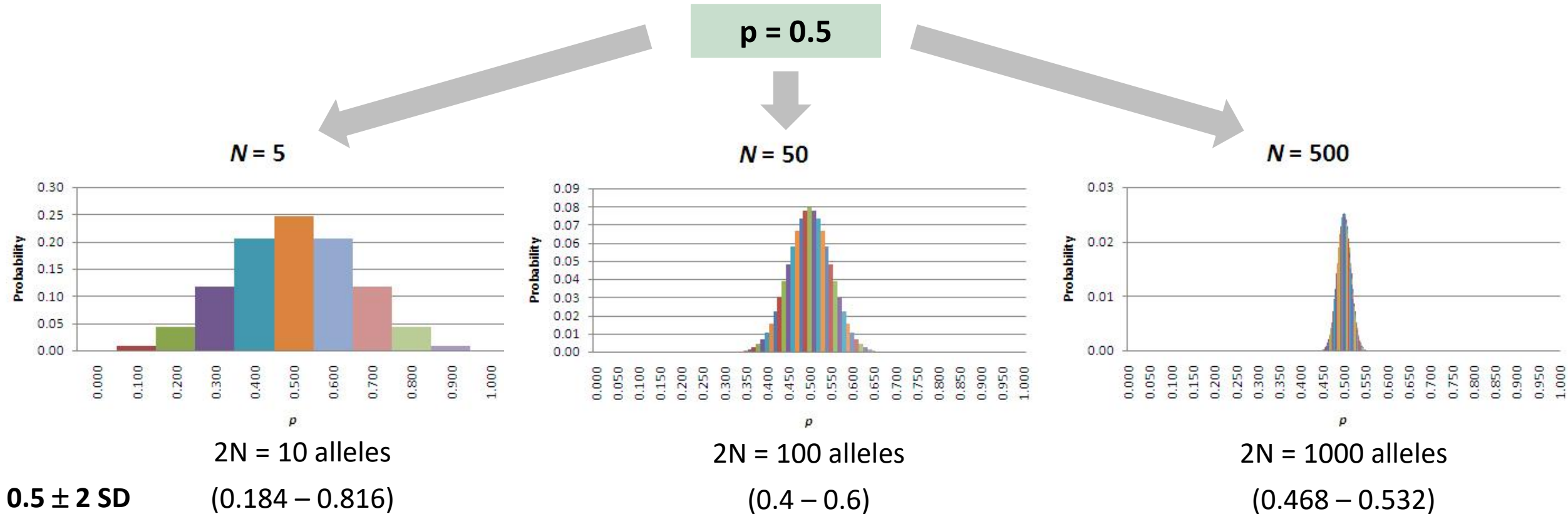
In a particular generation, the frequency of allele A_1 is 0.7 and the frequency of allele A_2 is 0.3.

Which is the probability of having allele frequencies 0.5 for each allele in the next generation?



Decrease in sampling error with increasing population size

The breadth of the distribution narrows as population size increases due to a decrease in sampling error

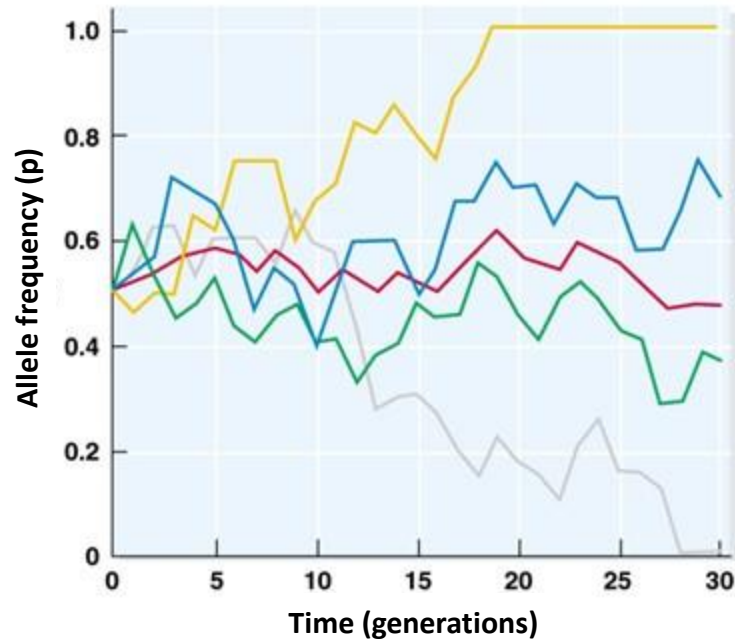


$0.5 \pm 2 \text{ SD}$
Confidence
interval 95%

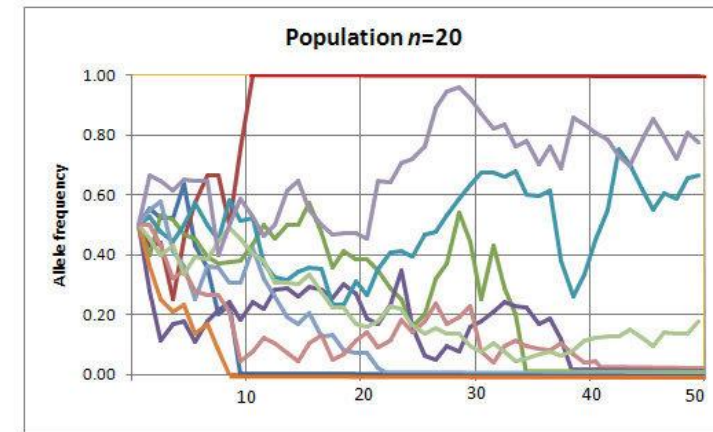
**Allele frequencies will change by chance in population of all sizes,
but the amount of change due to sampling error decreases as population size increases**

Allele frequencies over time in populations of different sizes

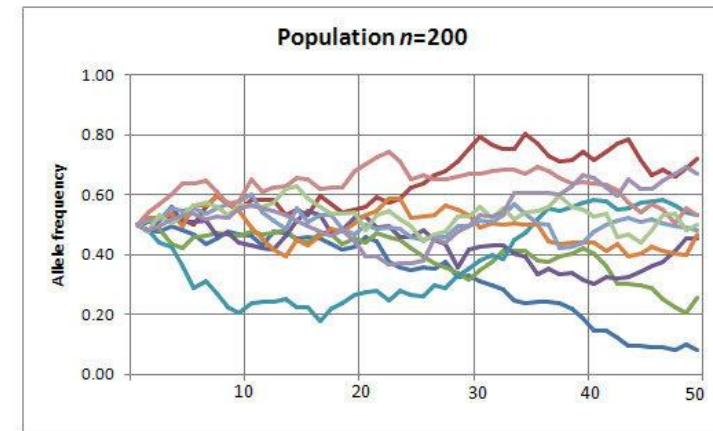
Graph explanation



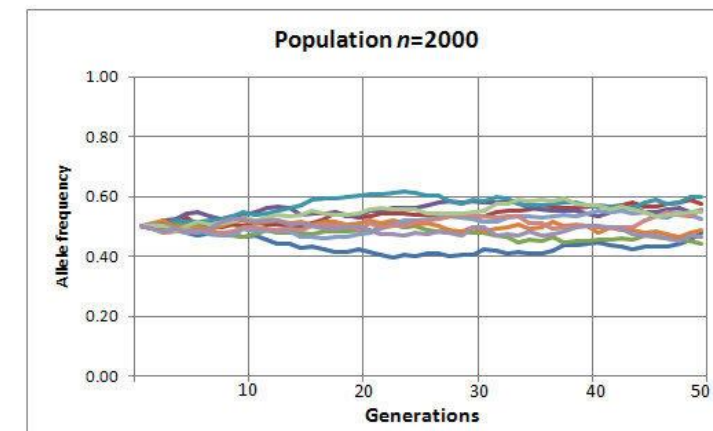
- Lines correspond to independent populations/replicates
- Random changes from one generation to the next
- Allele frequencies that reach upper or lower axis represent cases of fixation or loss



$2N = 40$



$2N = 400$

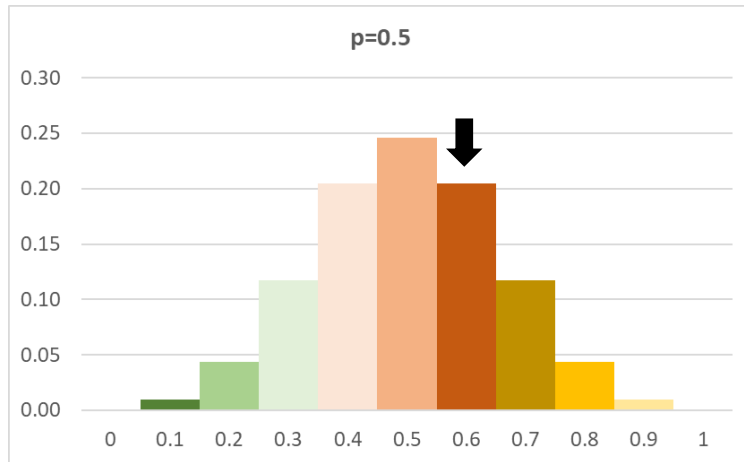


$2N = 4000$

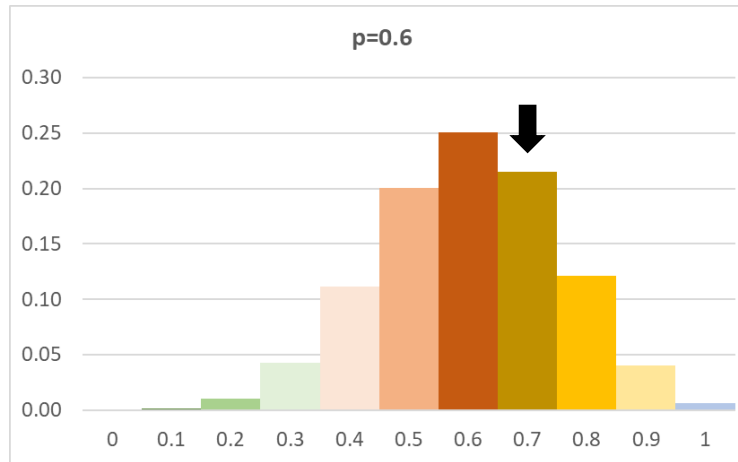
Sampling to produce next generation will center on the new value of p

$2N = 10$

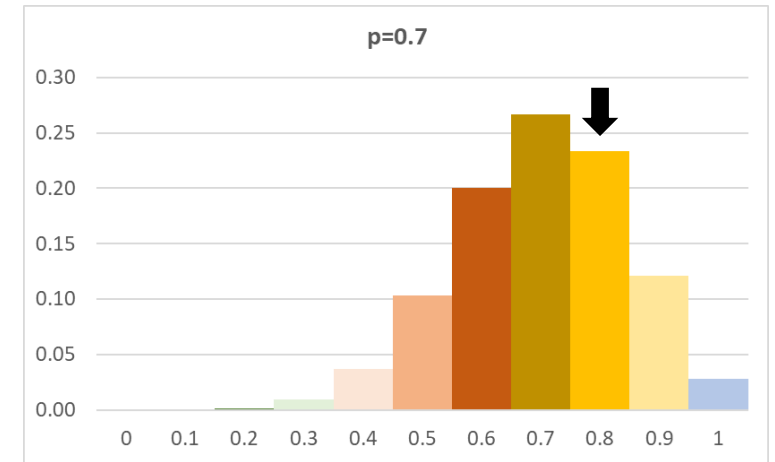
Generation 0



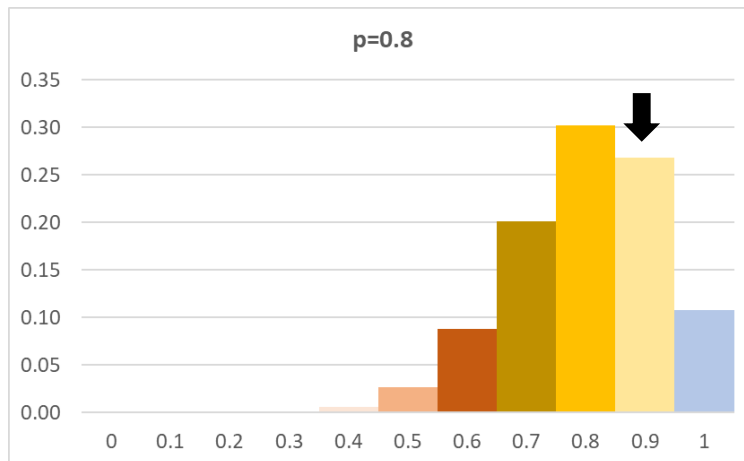
Generation 1



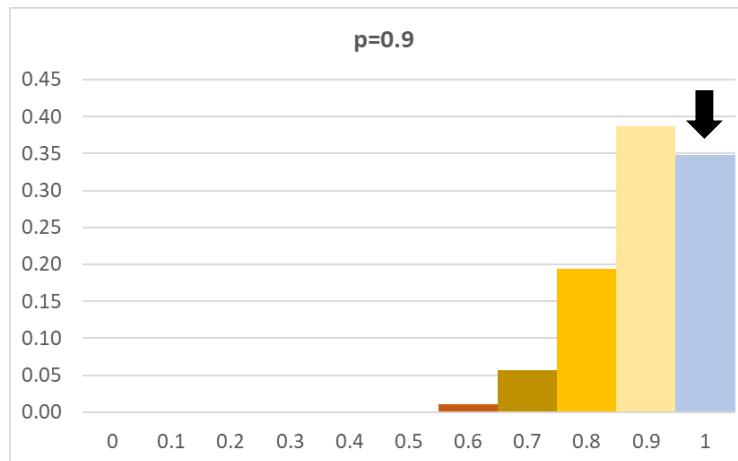
Generation 2



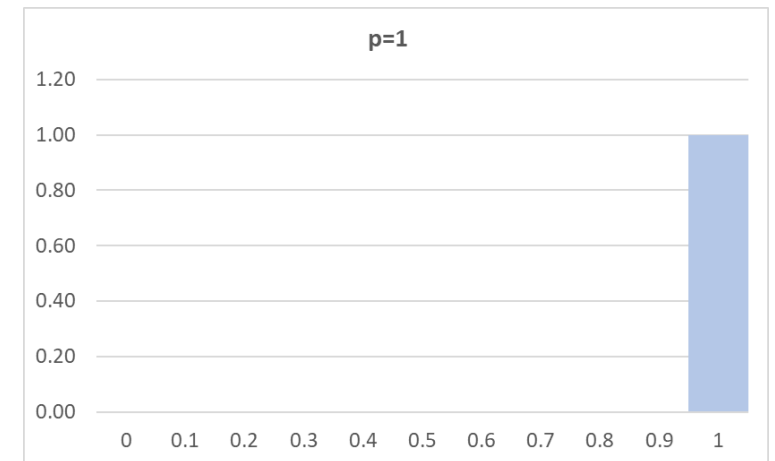
Generation 3



Generation 4



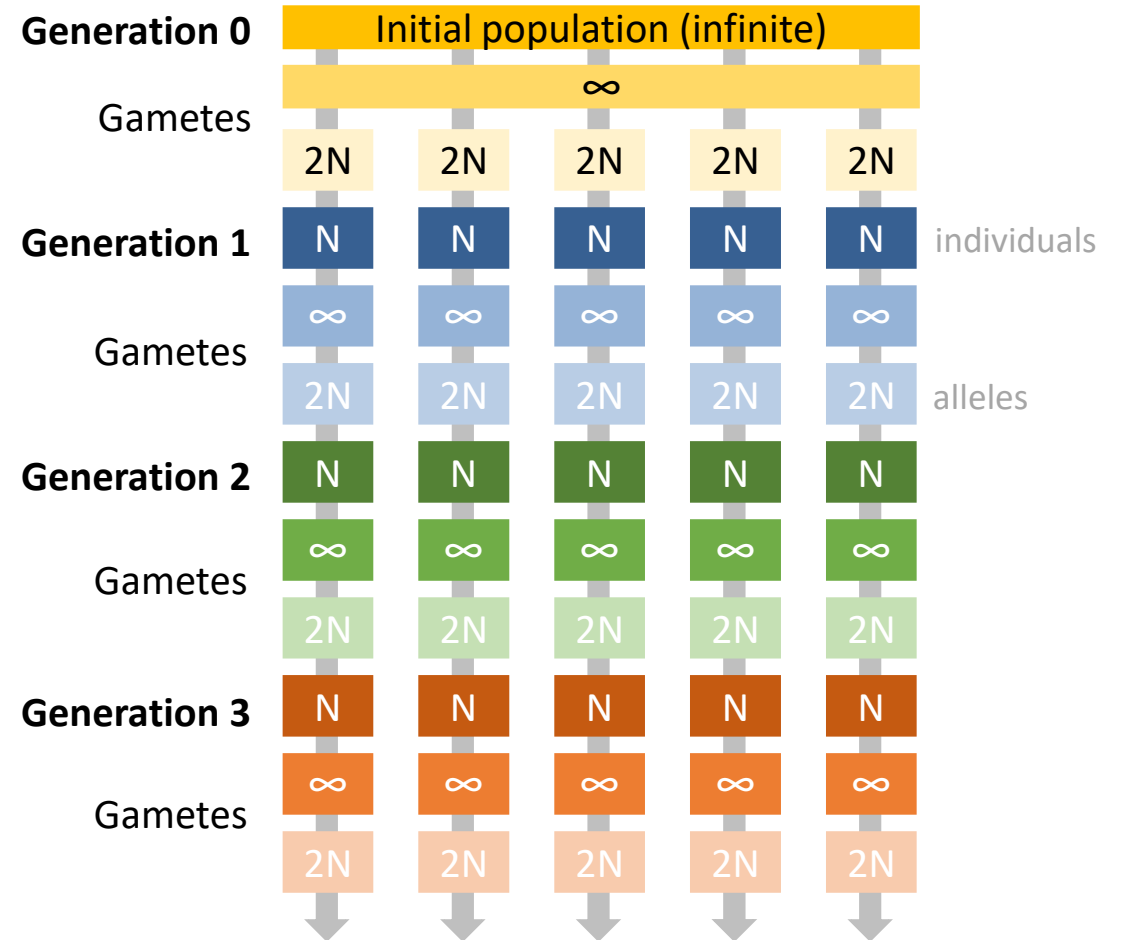
Generation 5



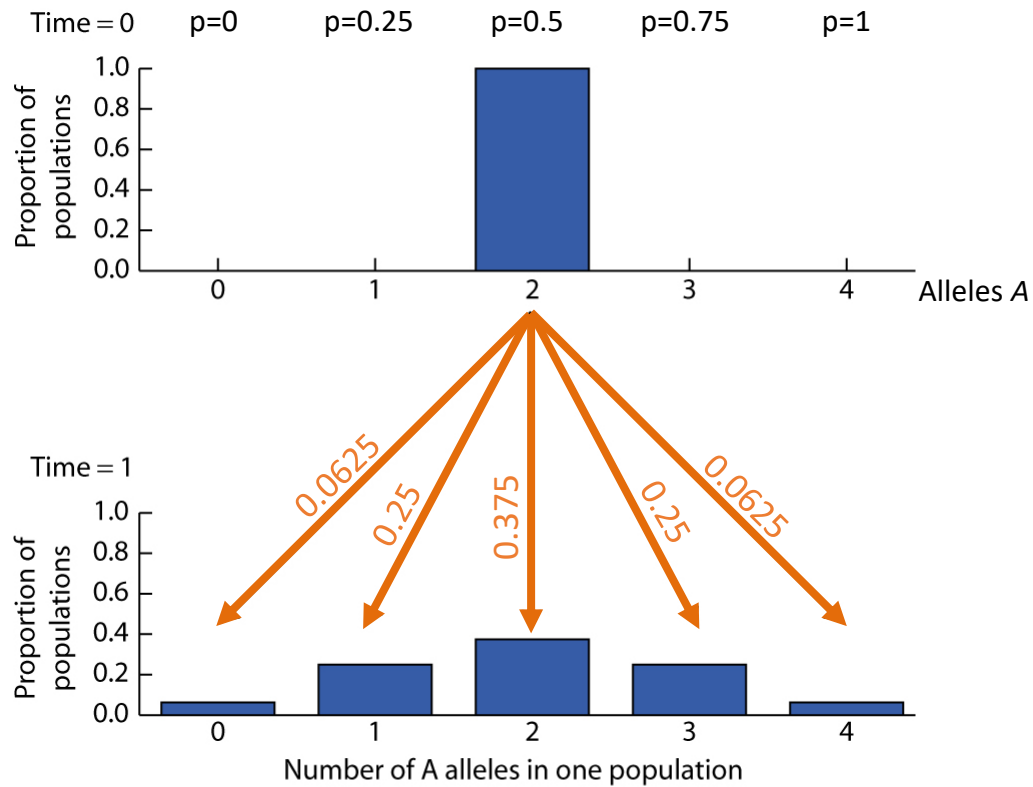
Wright-Fisher model for genetic drift

ASSUMPTIONS

1. Infinite populations
2. Constant size N
3. Random mating
4. Isolated populations (no migration)
5. No mutation
6. All individuals contribute equally to the infinite pool of gametes
7. Each generation is formed by a random sample of $2N$ gametes from the previous generation



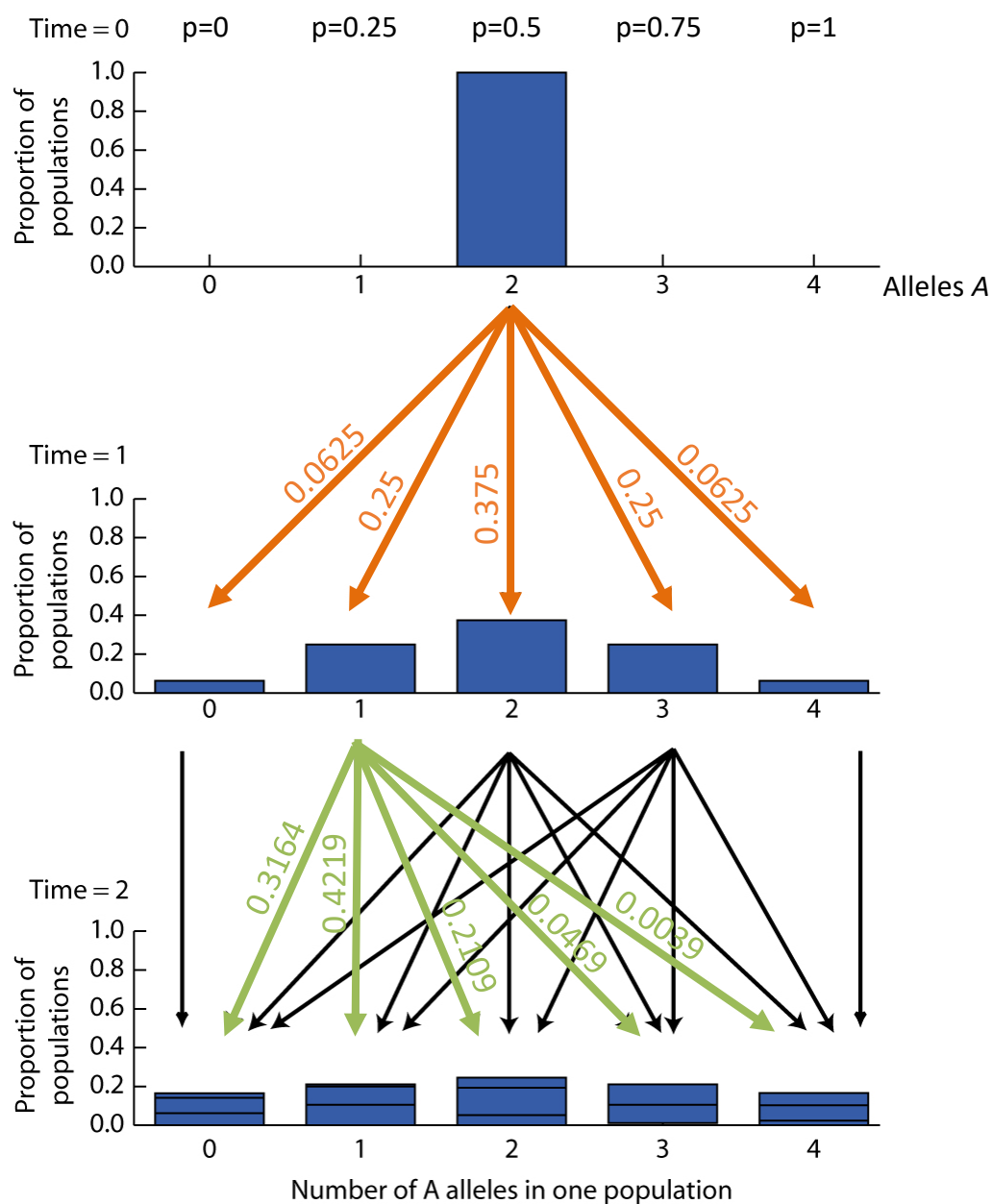
Markov chains



N = 2 individuals 2N = 4 alleles $p_0 = q_0 = 0.5$

A alleles	a alleles	Probability
0	4	$1 \cdot (0.5)^0 \cdot (0.5)^4 = 0.0625$
1	3	$\frac{4!}{1! 3!} \cdot (0.5)^1 \cdot (0.5)^3 = 0.25$
2	2	$\frac{4!}{2! 2!} \cdot (0.5)^2 \cdot (0.5)^2 = 0.375$
3	1	$\frac{4!}{3! 1!} \cdot (0.5)^3 \cdot (0.5)^1 = 0.25$
4	0	$1 \cdot (0.5)^4 \cdot (0.5)^0 = 0.0625$

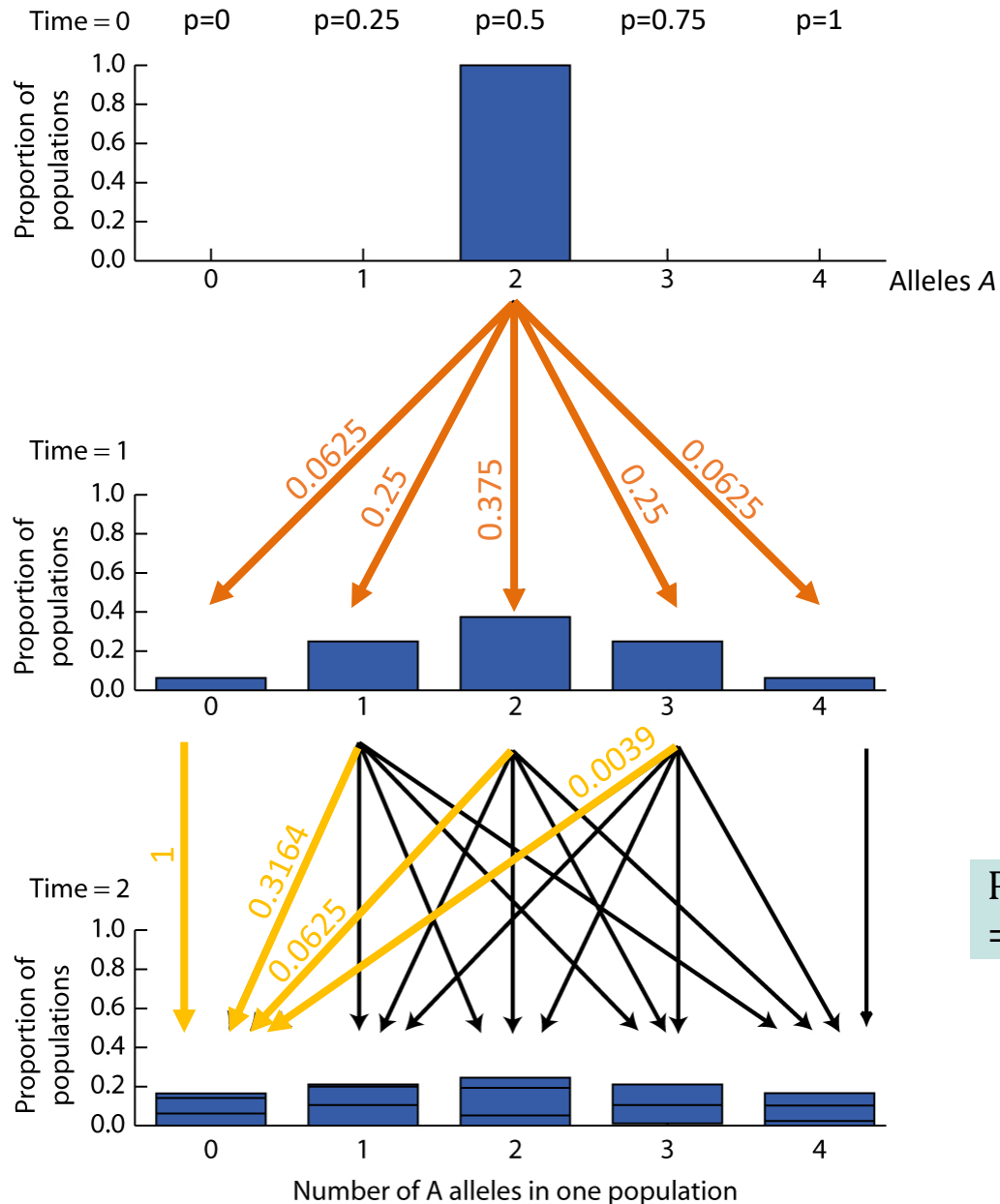
Markov chains



Probability transition matrix for a population of size $2N = 4$

		Number of alleles A in generation t				
		0	1	2	3	4
Number of alleles A in generation t+1	0	1	0.3164	0.0625	0.0039	0
	1	0	0.4219	0.25	0.0469	0
	2	0	0.2109	0.375	0.2109	0
	3	0	0.0469	0.25	0.4219	0
	4	0	0.0039	0.0625	0.3164	1

Markov chains



Probability transition matrix for a population of size $2N = 4$

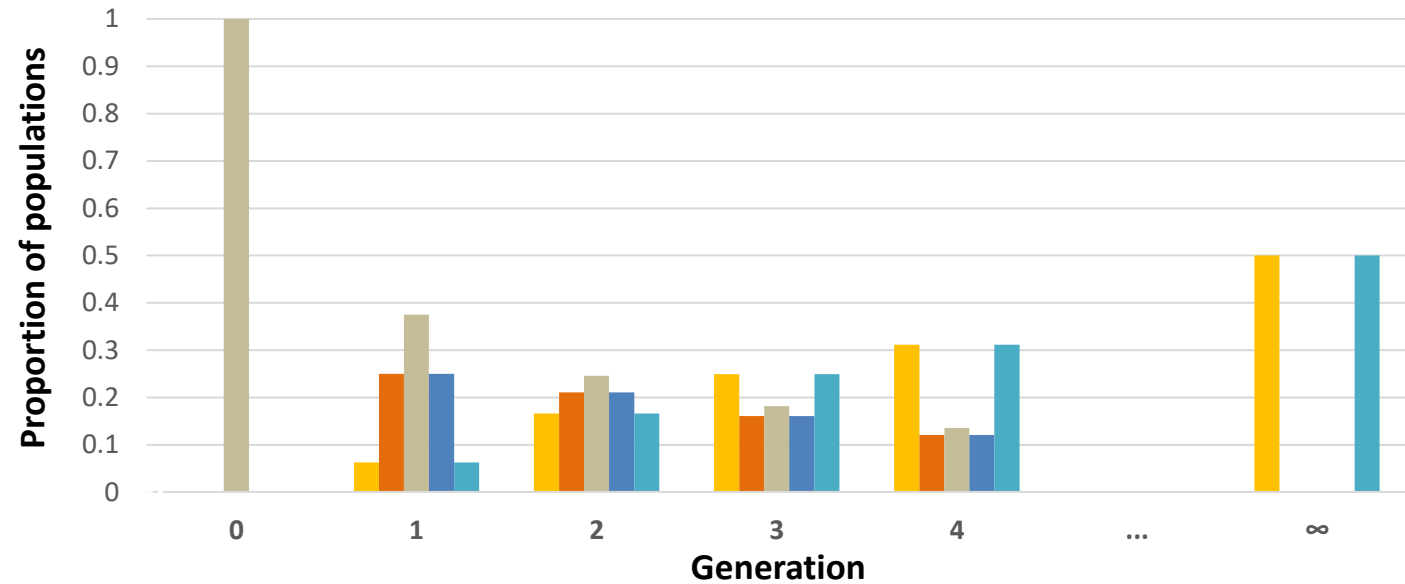
Number of alleles A in generation t+1	Number of alleles A in generation t					
	0	1	2	3	4	
	0	1	0.3164	0.0625	0.0039	0
	1	0	0.4219	0.25	0.0469	0
	2	0	0.2109	0.375	0.2109	0
	3	0	0.0469	0.25	0.4219	0
	4	0	0.0039	0.0625	0.3164	1

$$P(0 \text{ alleles A in generation 2}) = (0.0625 \cdot 1) + (0.25 \cdot 0.3164) + (0.375 \cdot 0.0625) + (0.25 \cdot 0.0039) = 0.166$$

Genetic drift will cause allele fixation

N = 2 individuals
2N = 4 alleles
p₀ = q₀ = 0.5

- An increasing number of populations accumulate at states of 0 and 4 alleles A, eventually reaching fixation or loss for all populations
- Mean frequency does not change with time
- Mean heterozygosity decreases with time
- Variance increases with time



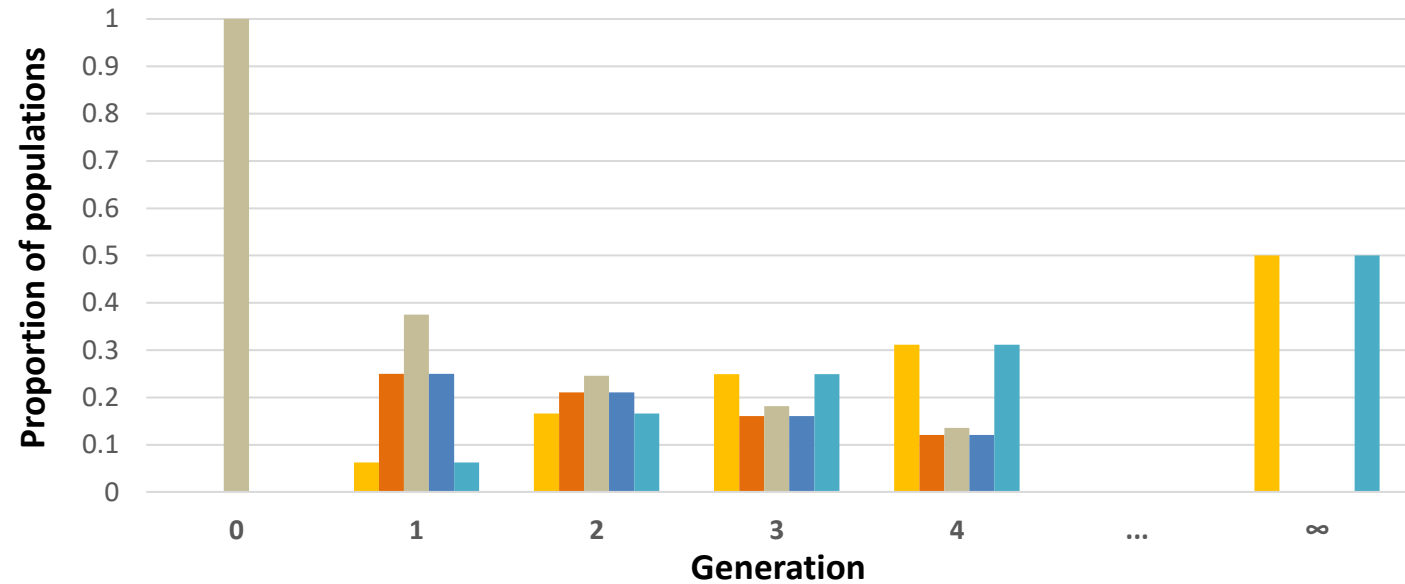
	0	1	2	3	4	...	∞
0 (p=0)	0	0.0625	0.1660	0.2490	0.3117	...	0.5
1 (p=0.25)	0	0.25	0.2109	0.1604	0.1205	...	0
2 (p=0.5)	1	0.375	0.2461	0.1813	0.1356	...	0
3 (p=0.25)	0	0.25	0.2109	0.1604	0.1205	...	0
4 (p=1)	0	0.0625	0.1660	0.2490	0.3117	...	0.5
\bar{p}	0.5	0.5	0.5	0.5	0.5	...	0.5
\bar{H}	0.5	0.375	0.2812	0.2109	0.1582	...	0
Var(p)	0	0.0625	0.1094	0.1445	0.1709	...	0.25

Genetic drift will cause allele fixation

Fixation index

The fixation index (F_{ST}) measures how far a group of populations is into the genetic drift process (which ends in the equilibrium with all populations with a fixed allele)

$$F_{ST} = \frac{\text{Var}(t)}{\text{Var}(\infty)} = 1 - \left(1 - \frac{1}{2N}\right)^t$$



	0	1	2	3	4	...	∞
0 (p=0)	0	0.0625	0.1660	0.2490	0.3117	...	0.5
1 (p=0.25)	0	0.25	0.2109	0.1604	0.1205	...	0
2 (p=0.5)	1	0.375	0.2461	0.1813	0.1356	...	0
3 (p=0.25)	0	0.25	0.2109	0.1604	0.1205	...	0
4 (p=1)	0	0.0625	0.1660	0.2490	0.3117	...	0.5
\bar{p}	0.5	0.5	0.5	0.5	0.5	...	0.5
\bar{H}	0.5	0.375	0.2812	0.2109	0.1582	...	0
Var(p)	0	0.0625	0.1094	0.1445	0.1709	...	0.25

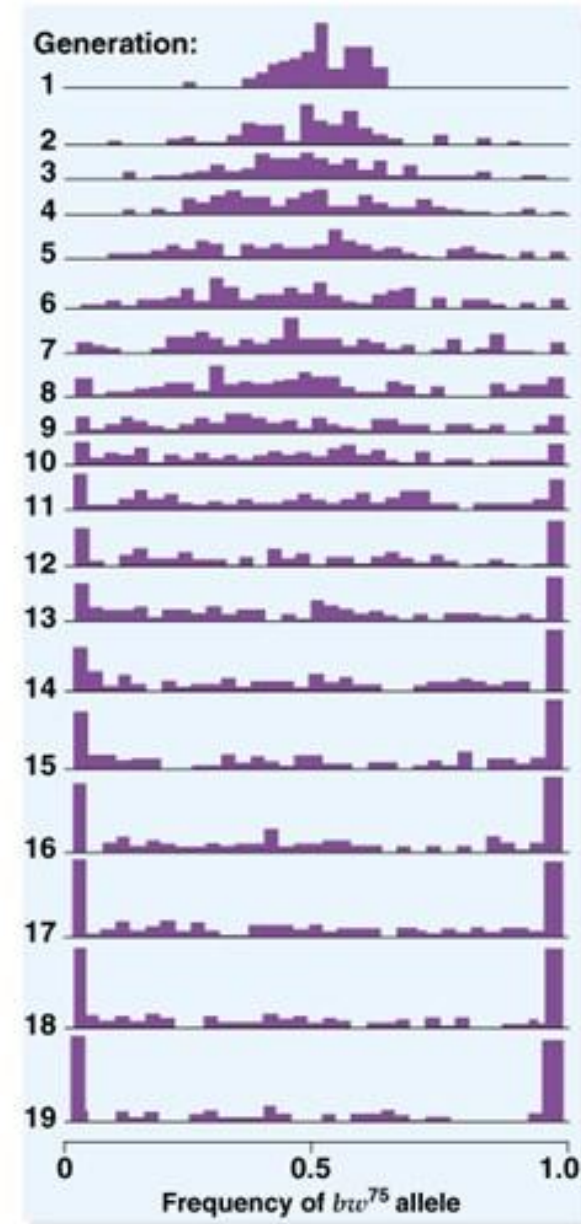
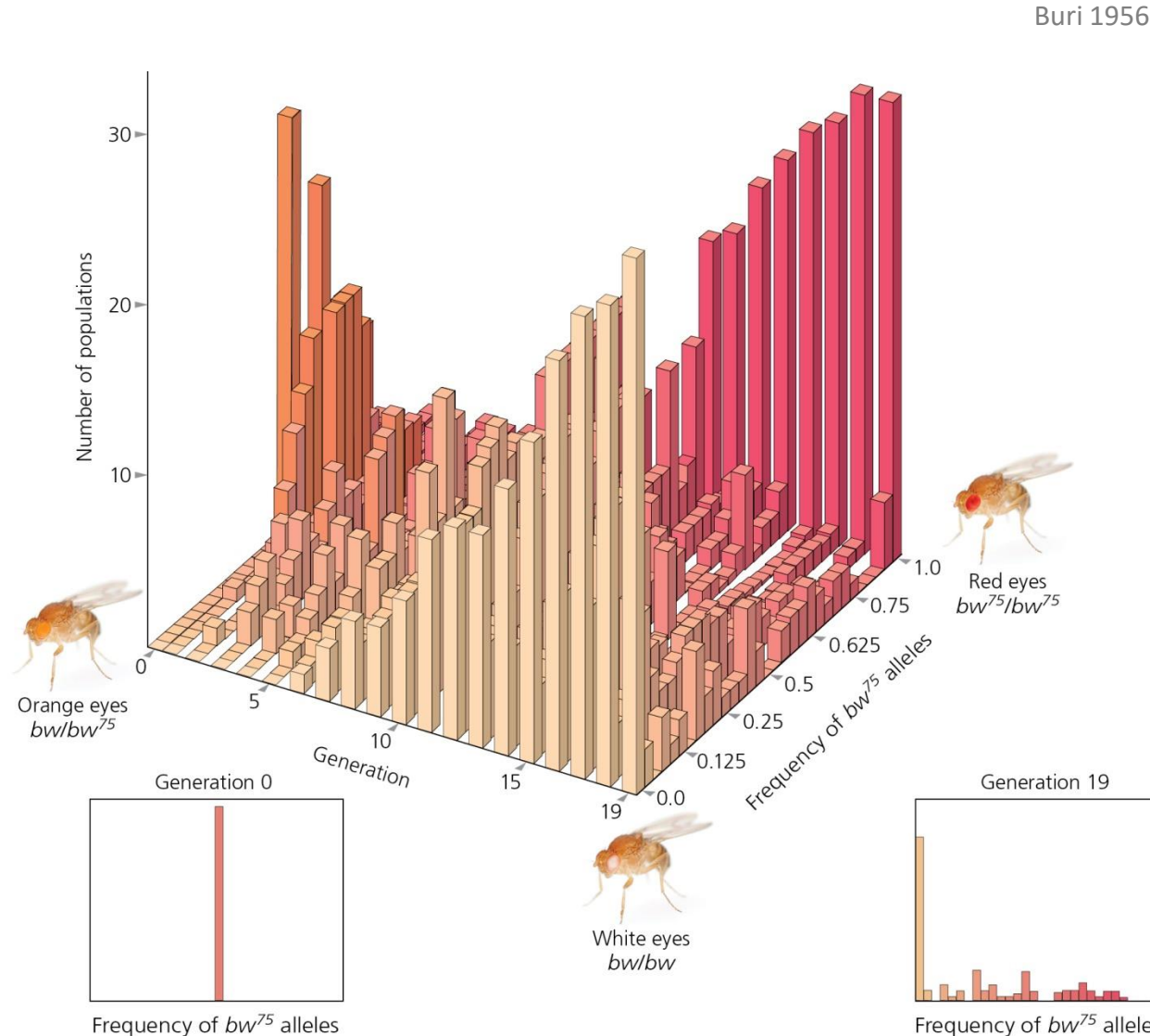
Genetic drift in *Drosophila* experiment

EXPERIMENT

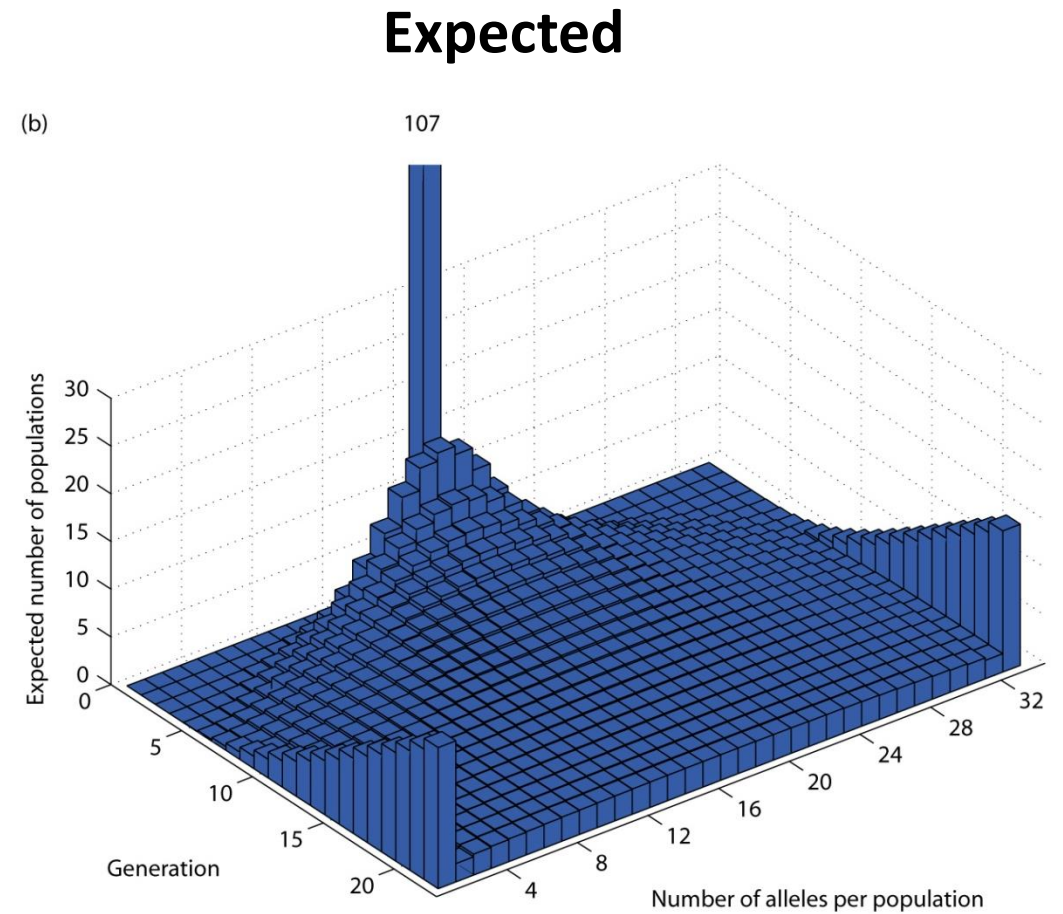
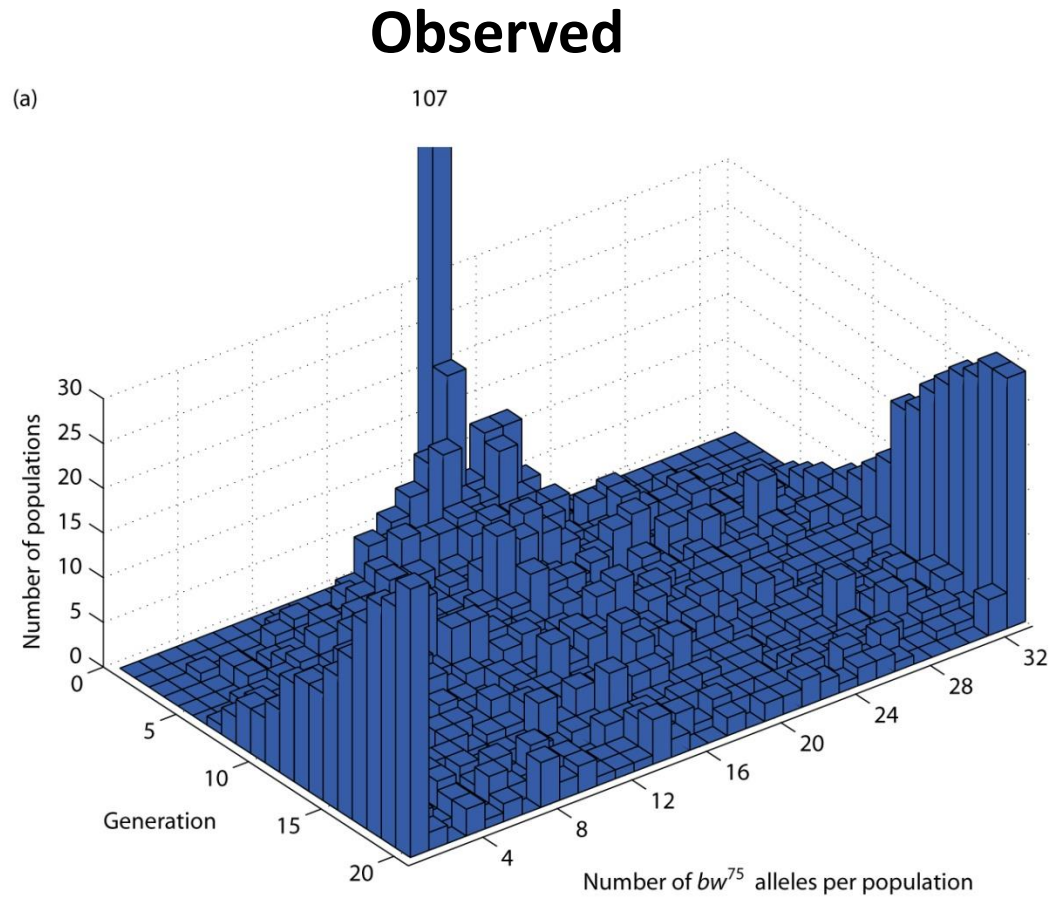
- 107 *D. melanogaster* populations
- $p_0 = 0.5$ (bw^{75}/bw alleles)
- $N=16$ (8 ♂ + 8 ♀ chosen at random each generation)
- 19 generations

RESULTS

- Populations fixed for bw or bw^{75} alleles increase over time
- 50% of populations fixed for each allele
- Flat distribution among unfixed populations

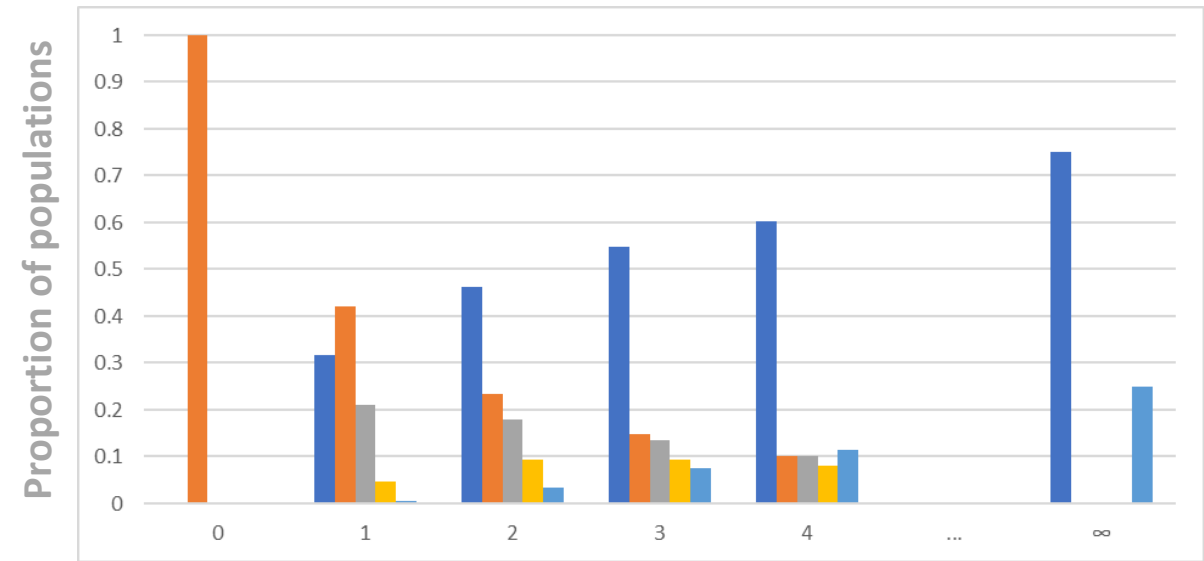


Wright-Fisher model for genetic drift

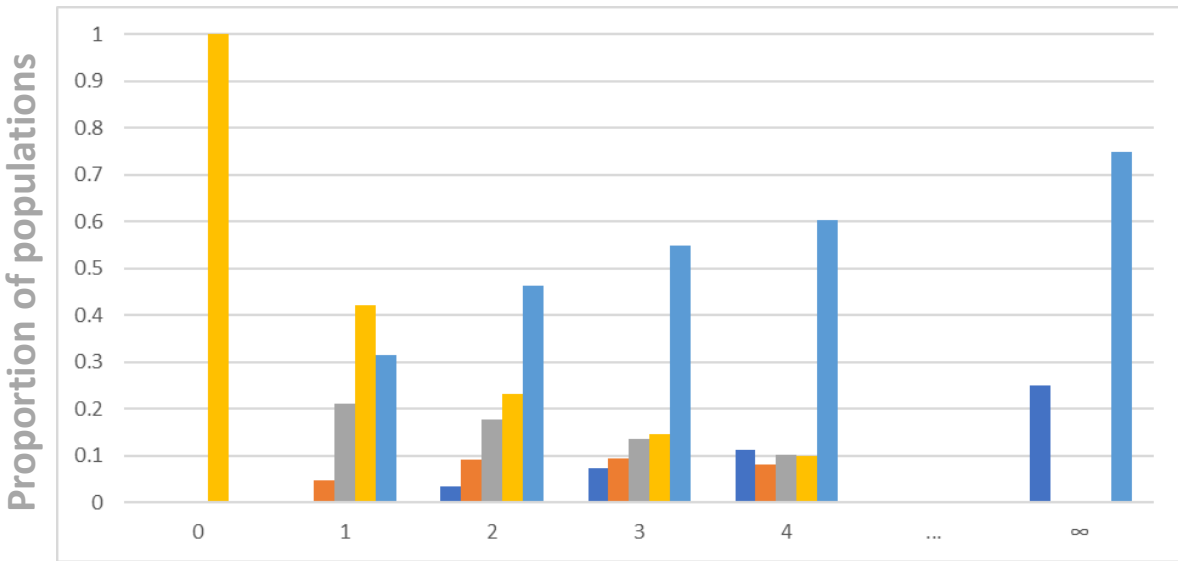


Probability of fixation of a neutral allele

$p_0 = 0.25$



$p_0 = 0.75$



A alleles	0	1	2	3	4	...	∞
0 ($p=0$)	0	0.3164	0.4633	0.5484	0.6038	...	0.75
1 ($p=0.25$)	1	0.4219	0.2329	0.1471	0.1003	...	0
2 ($p=0.5$)	0	0.2109	0.1780	0.1353	0.1017	...	0
3 ($p=0.75$)	0	0.0469	0.0923	0.0944	0.0805	...	0
4 ($p=1$)	0	0.0039	0.0336	0.0748	0.1137	...	0.25

A alleles	0	1	2	3	4	...	∞
0 ($p=0$)	0	0.0039	0.0336	0.0748	0.1137	...	0.25
1 ($p=0.25$)	0	0.0469	0.0923	0.0944	0.0805	...	0
2 ($p=0.5$)	0	0.2109	0.1780	0.1353	0.1017	...	0
3 ($p=0.75$)	1	0.4219	0.2329	0.1471	0.1003	...	0
4 ($p=1$)	0	0.3164	0.4633	0.5484	0.6038	...	0.75

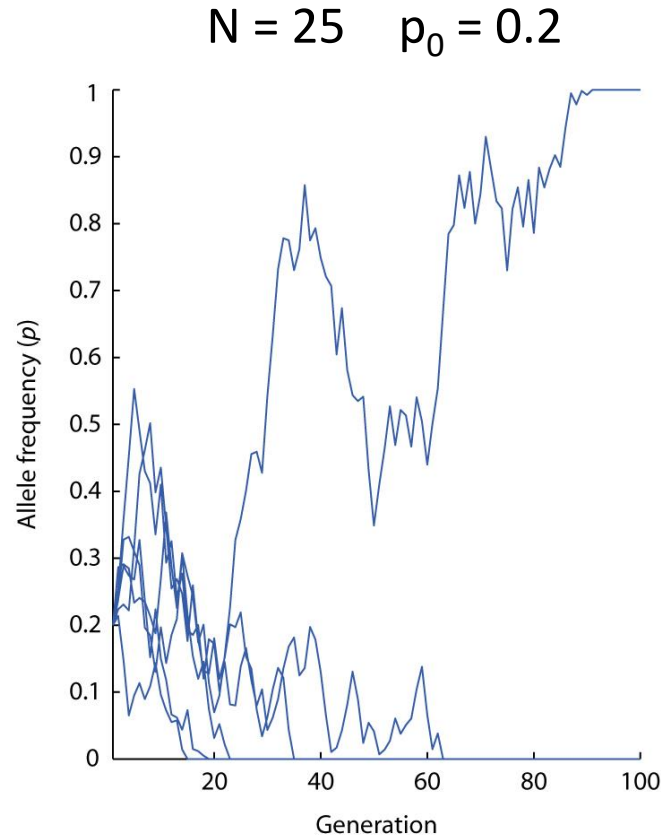
Probability of fixation of a neutral allele is equal to its initial frequency in the population

Probability of fixation of a neutral allele

Probability of fixation (P_{fix}) of a neutral allele is equal to its initial frequency in the population

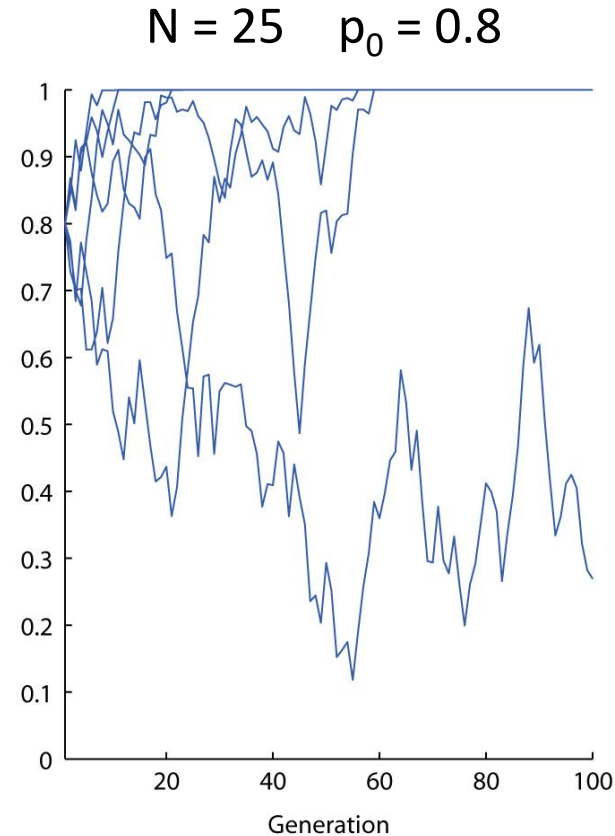
More populations go
to loss of the allele

$$P_{\text{fix}} = 0.2$$



More populations
reach fixation

$$P_{\text{fix}} = 0.8$$



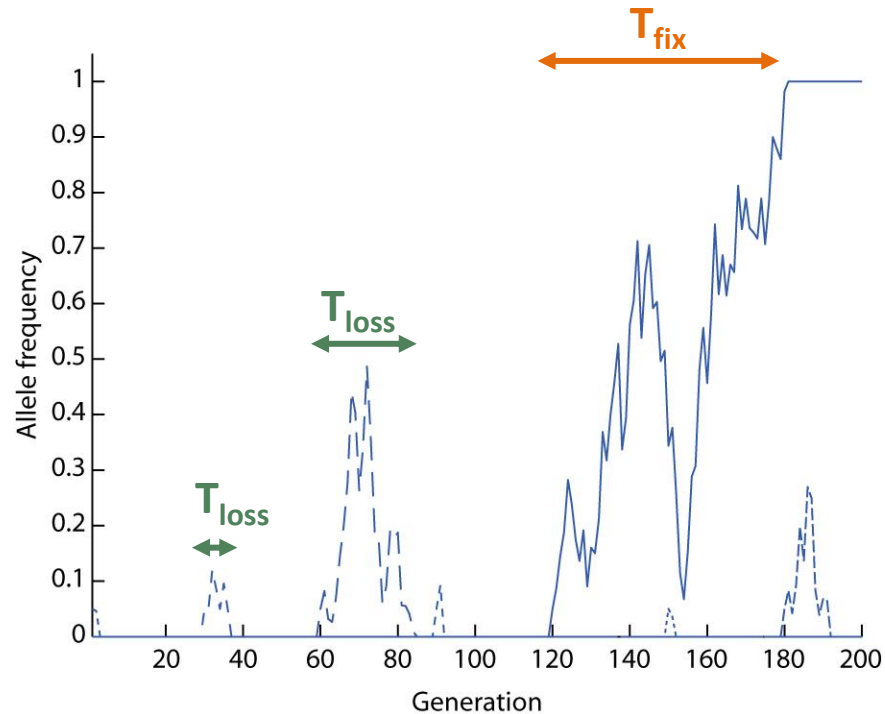
$$P_{\text{fix}} = p_0$$

A new allele with $p_0 = \frac{1}{2N}$ is more likely to be lost than fixed

Average time to fixation or loss of a new allele

The average time that an allele takes to reach fixation or loss depends on its initial frequency when under the influence of genetic drift alone

New alleles introduced every 30 generations into a population of $N_e = 10$



NEW ALLELE $p = \frac{1}{2N}$

$$P_{fix} = p_0$$

Average time to fixation

$$T_{fix} \approx 4N$$

Average time to loss

$$T_{loss} \approx 2 \ln(2N)$$

New allele $N = 1000$

$$P_{fix} = \frac{1}{2000} = 0.0005 \quad T_{fix} = 4000 \text{ generations} \quad T_{loss} = 15 \text{ generations}$$

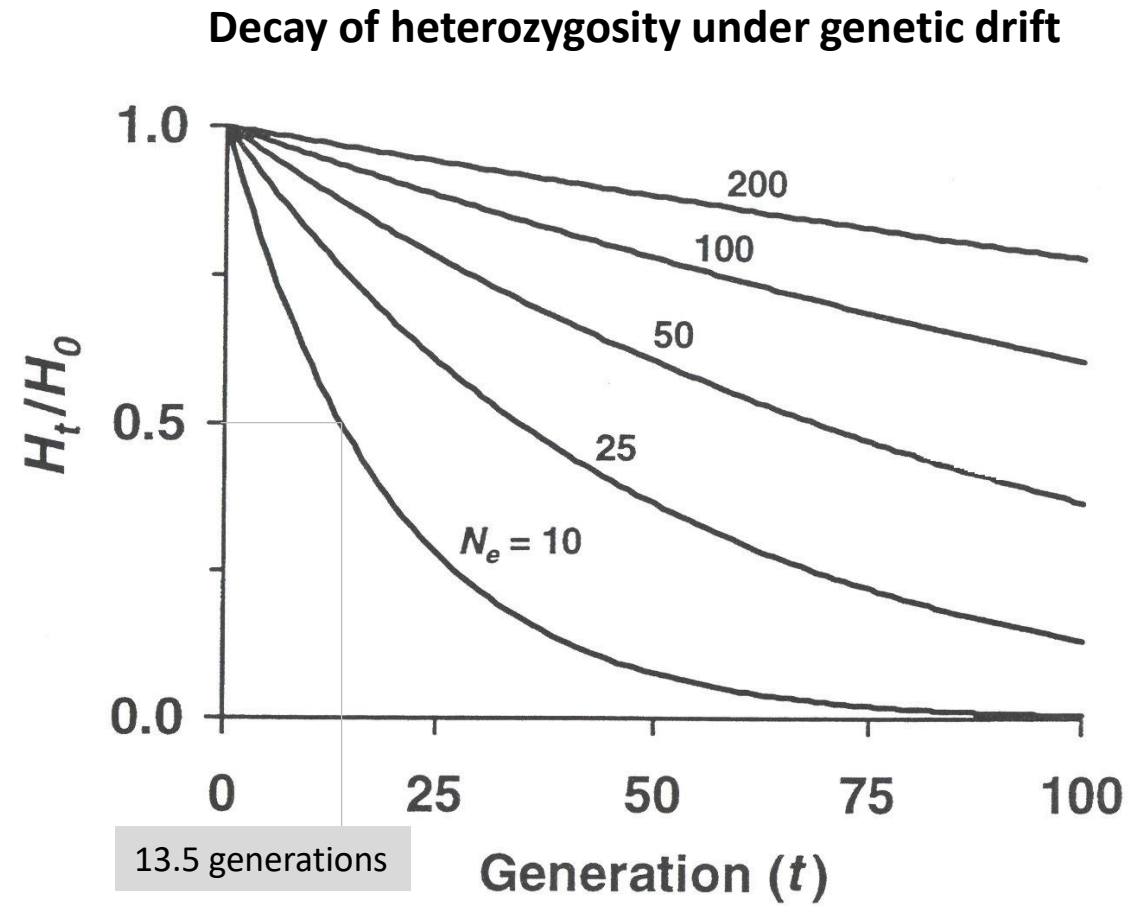
Genetic drift causes a reduction in heterozygosity

Heterozygosity = observed proportion of heterozygotes in a population

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

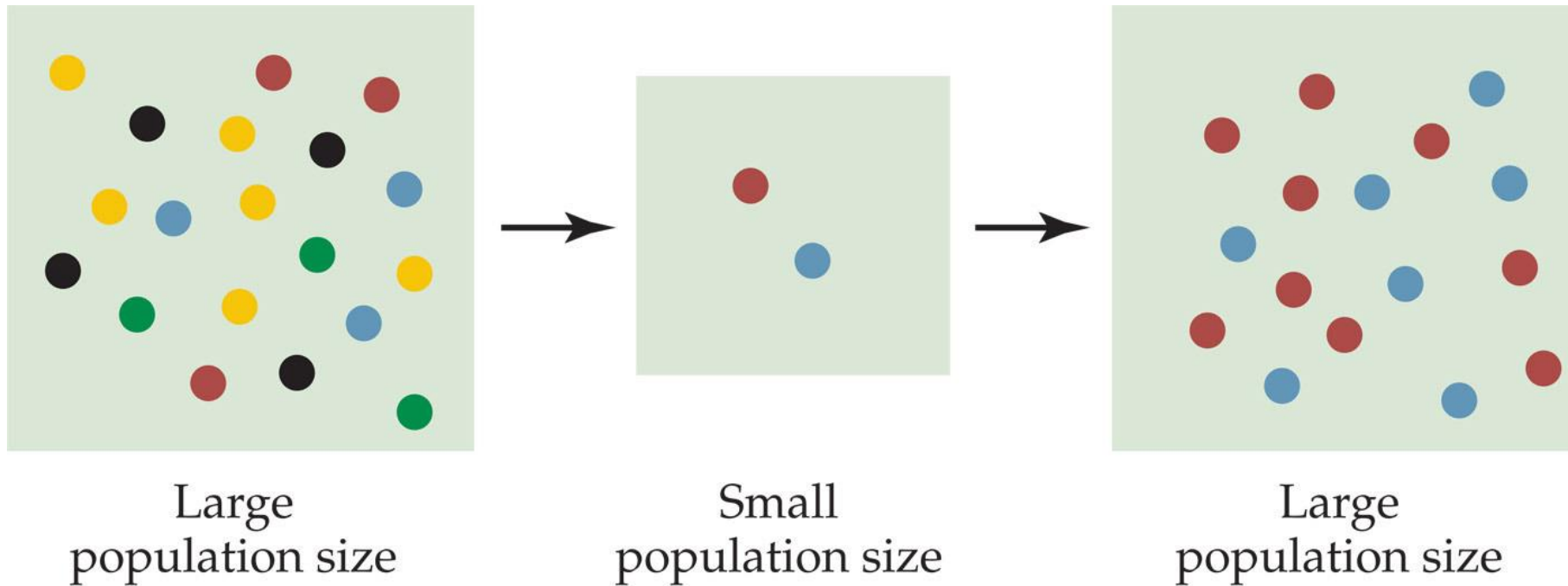
Heterozygosity declines by a factor of $1 - \frac{1}{2N}$ every generation due to drift

The smaller the population, the stronger the effect of drift, and the faster that heterozygosity will be lost



When one allele is fixed, $H = 0$

Reductions in population size



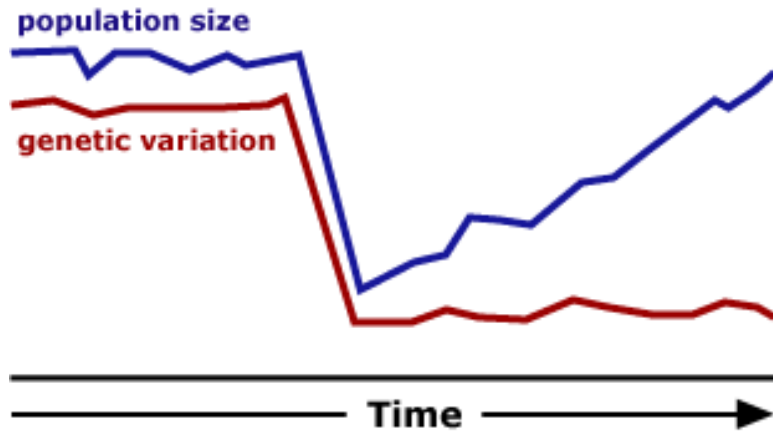
Genetic drift acts more quickly to reduce genetic variation in small populations

The resulting population can have:

1. Reduced variation and reduced ability to adapt to new selection pressures
2. A non-random sample of the genes in the original population

Reductions in population size

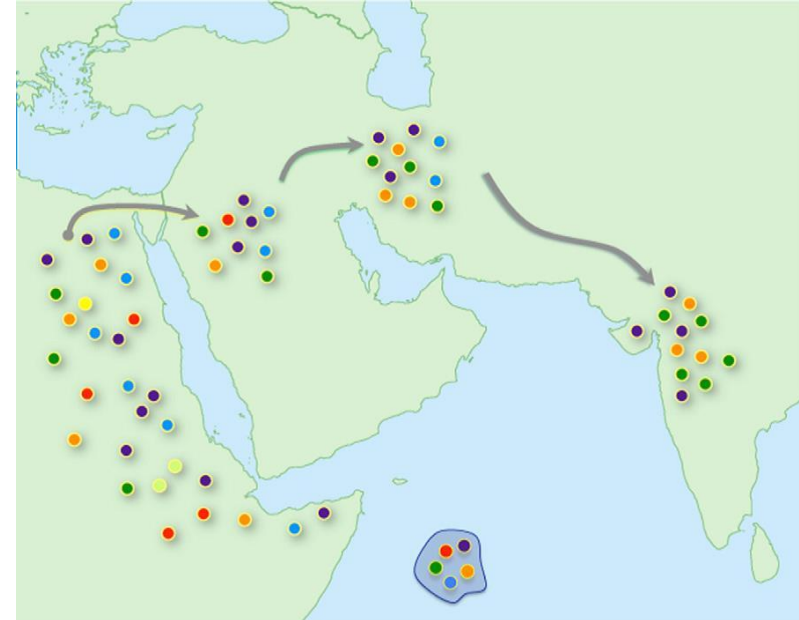
Bottleneck



A **bottleneck** occurs when a population's size is reduced for at least one generation.

Consequences: long-term reduction of genetic variation
(even if the bottleneck does not last for many generations and the population regains its previous size)

Founder effect

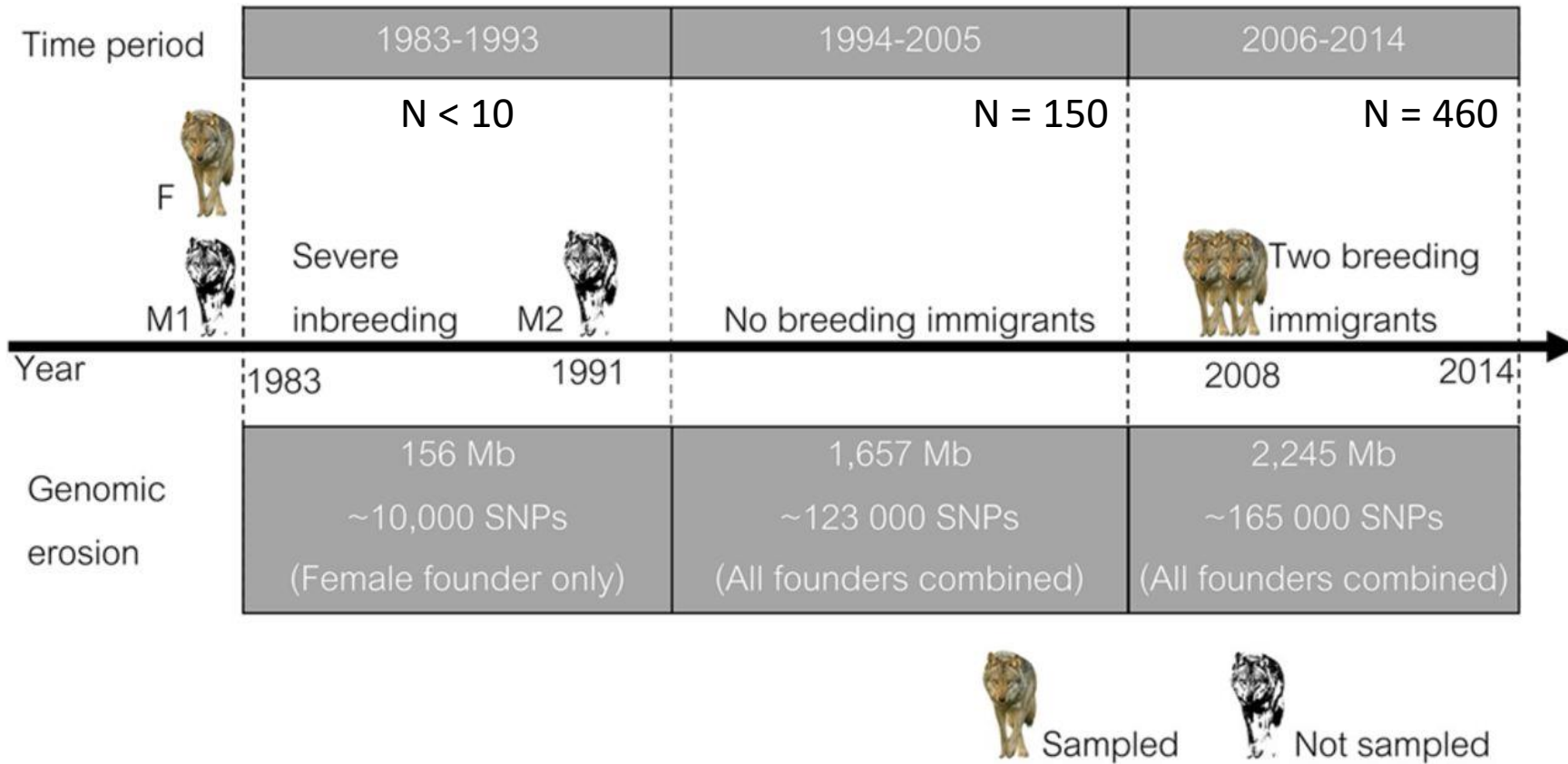


A **founder effect** occurs when a new colony is started by a few members of the original population.

Consequences:

1. Substantial loss in genetic diversity
2. Fast divergence between source and founder populations

Consequences of genetic drift on a small population

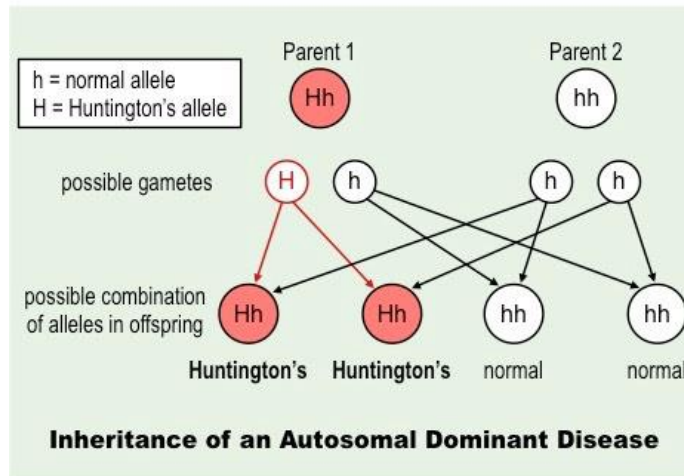


- Scandinavian gray wolf (*Canis lupus*) extinction in the late 1960s by human persecution
- Population reestablished by 3 immigrant founders (1 F + 2 M)
- Sequencing data of 76 Scandinavian wolves sampled over a period of 30 years were used to reconstruct chromosome-level haplotypes of the founders
- 10%–24% of their diploid genomes had become lost after about 20 years (\approx 5 generations) of drift
- Lost SNPs may include variants that are advantageous

Reduced genetic diversity levels due to:

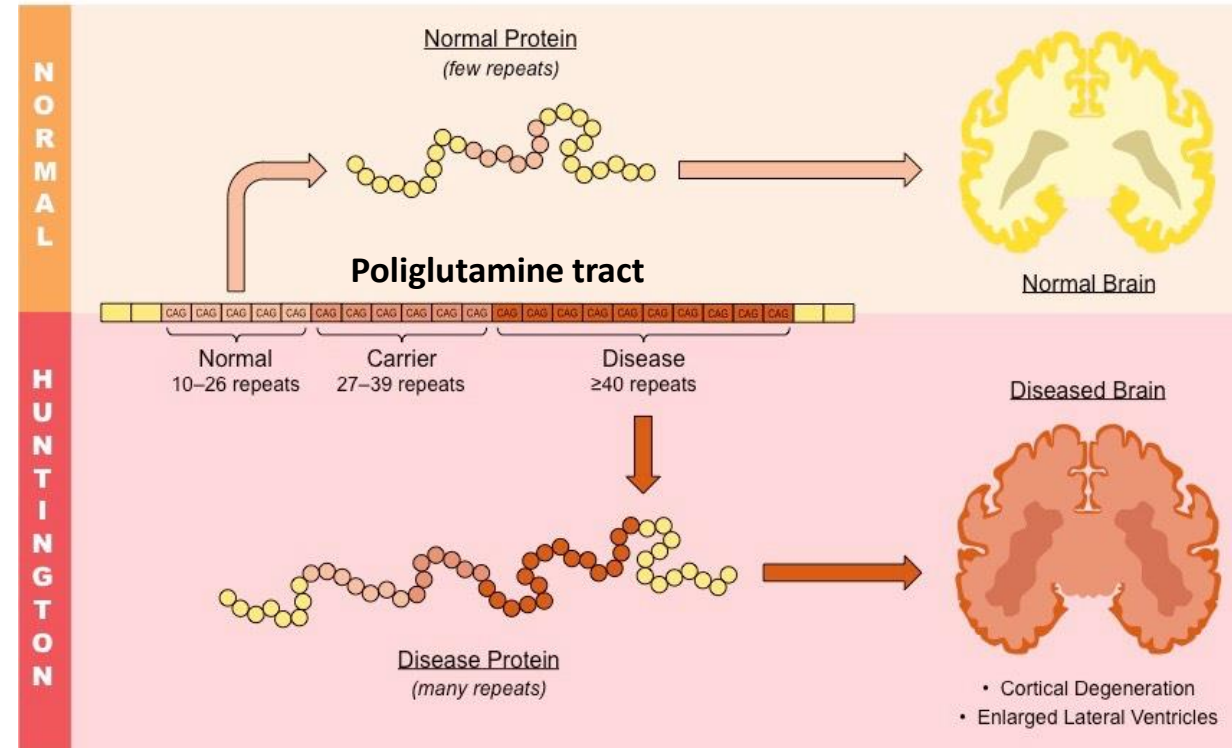
1. Limited input (only 3 founders) and absence of gene flow from neighboring populations until 2008
2. Severe inbreeding and drift reduced further the limited diversity

Founder effect – Huntington's disease in Venezuela



In the fishing villages located near Lake Maracaibo in Venezuela, there are more people with Huntington's disease than anywhere else in the world. In some villages, more than half the people may develop the disease. Usual incidence is 30-70 cases per million people in most Western countries.

Huntington's disease is inherited as an autosomal dominant trait and causes degeneration of nerve cells in the brain ultimately leading to death.



Founder effect

+

Mutation

+

Weak selection

In the 1800s a single woman with Huntington's allele had 10 children and today many individuals trace their ancestry back to this lineage

Triplet expansion creates new disease alleles

Disease does not affect people until *after* they've reproduced

Effective population size (N_e)

Ideal population ($N_c = N_e$)

1. There are equal numbers of males and females, all of whom are able to reproduce.
2. All individuals are equally likely to produce offspring, and the number of offspring that each produces varies no more than expected by chance.
3. Mating is random.
4. The number of breeding individuals is constant from one generation to the next.



**Most deviations will decrease
the effective population size**

Census population (N_c)

Total number of individuals in a population

Effective population size (N_e)

Individuals that actively participate in the reproductive process

Size of an idealized population that would have the same effect of random sampling on allele frequencies as that of the actual population

Effective population size (N_e)

N_e is usually much smaller than N_c

Factors that can contribute to this difference:

- 1. Different number of males and females**
- 2. Fluctuations in population size**
- 3. Variation in the number of offspring among individuals**
- 4. Bottlenecks**
- 5. Overlapping generations**

Different number of males and females

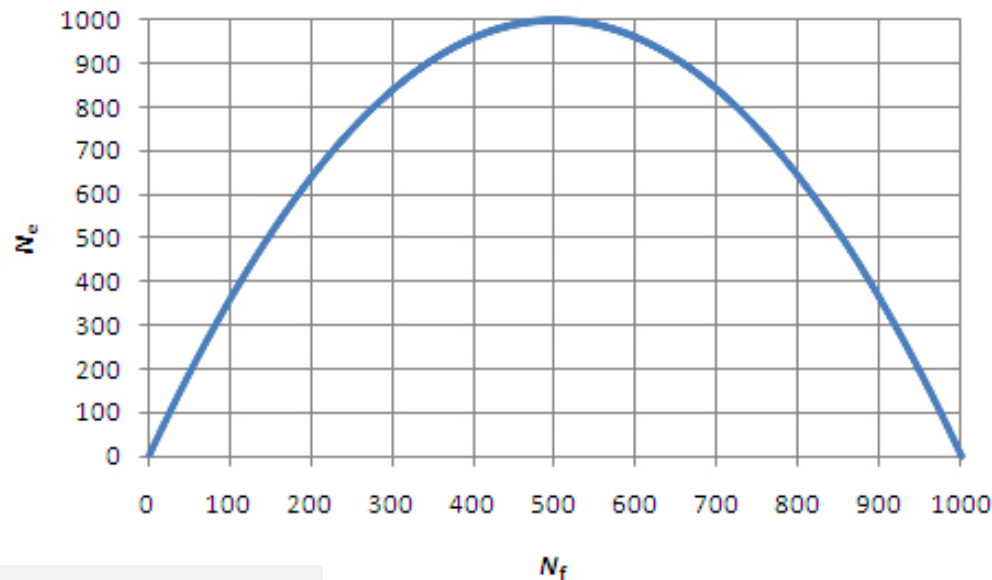
N_e in a population unequal sex ratio for autosomic genes

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

N_m = number of males
 N_f = number of females

Autosomal genes

Relationship between N_e and N_f in a population of 1000 mating individuals



What is the effective population size of a honey bee hive?

$N \approx 100.000$

$N_f = 1$

$N_m \gg \gg 1$

$N_e = 4$



Fluctuations in population size

Populations can show regular cycles of increase and decrease spanning a number of years.

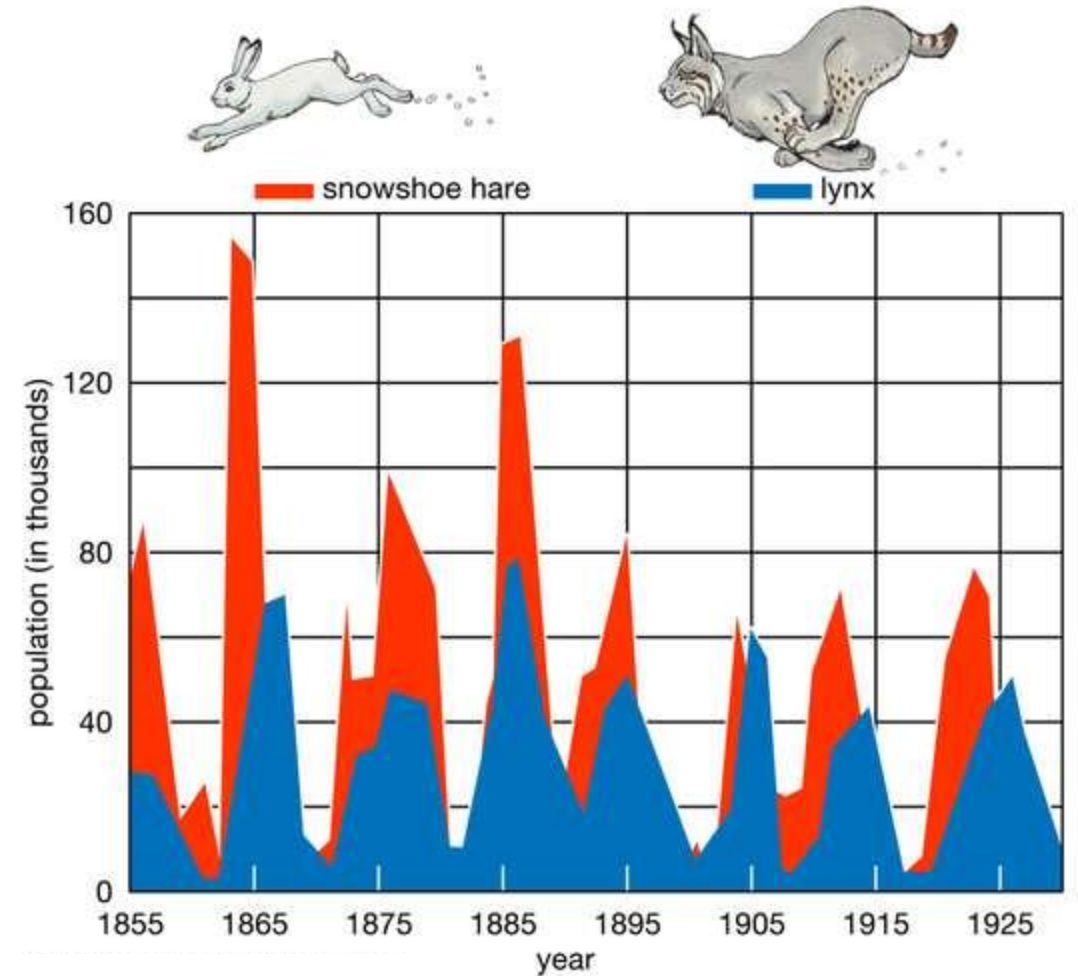
Small population numbers will cause an increased chance of fixation or loss of alleles by genetic drift

We can estimate the effect of fluctuations in populations on the overall effective size using the **harmonic mean**, which gives more weight to small values

$$\frac{1}{N_e} = \frac{1}{t} \left(\frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N_t} \right)$$

What is the effective population size of a population with 100 individuals that was reduced to 10 for 1 generation but has recovered its original census in the following generation?

$$N_e = 25$$



Cyclical fluctuations in the population density of the snowshoe hare and its effect on the population of its predator, the lynx.

Effective population sizes (N_e) in different species

Table 1 | **Effective population size (N_e) estimates from DNA sequence diversities**

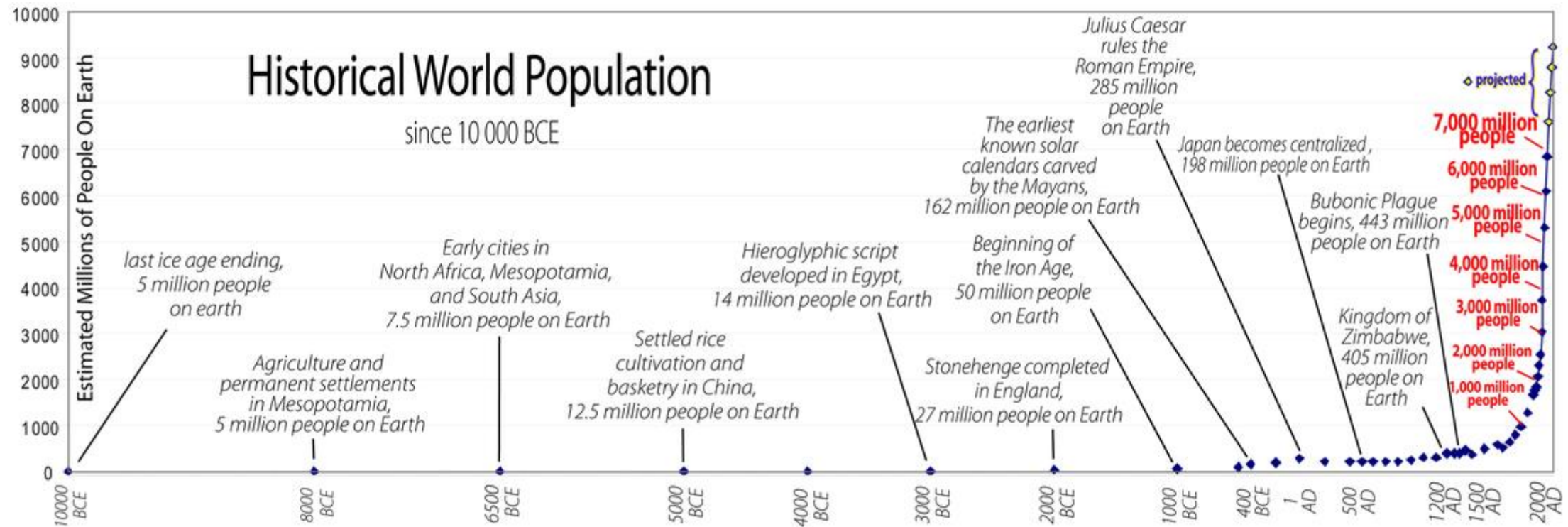
Species	N_e	Genes used	Refs
<i>Species with direct mutation rate estimates</i>			
Humans	10,400	50 nuclear sequences	145
<i>Drosophila melanogaster</i> (African populations)	1,150,000	252 nuclear genes	108
<i>Caenorhabditis elegans</i> (self-fertilizing hermaphrodite)	80,000	6 nuclear genes	41
<i>Escherichia coli</i>	25,000,000	410 genes	146
<i>Species with indirect mutation rate estimates</i>			
Bonobo	12,300	50 nuclear sequences	145
Chimpanzee	21,300	50 nuclear sequences	145
Gorilla	25,200	50 nuclear sequences	145
Gray whale	34,410	9 nuclear gene introns	147
<i>Caenorhabditis remanei</i> (separate sexes)	1,600,000	6 nuclear genes	43
<i>Plasmodium falciparum</i>	210,000 – 300,000	204 nuclear genes	148

For data from genes, synonymous site diversity for nuclear genes was used as the basis for the calculation, unless otherwise stated.

Low effective population sizes (N_e) in humans

The calculated N_e corresponds to the **historic size of human population**.

Even though the human population has exploded, our standing genetic variation largely reflects a much smaller past population size.



Genetic drift – Important ideas

- **Allele frequencies change randomly due to sampling error**
- **The direction of the change is unpredictable** (allele frequencies will randomly increase and decrease over time)
- **Cumulative behavior** (each generation allele frequency will tend to deviate more and more from initial frequency and probability of fixation increases with time)
- **The amount of change due to sampling error decreases as the population size increases** (smaller populations will be more affected by genetic drift than larger populations)
- Given enough time and in the absence of factors that maintain both alleles, **one allele will drift to fixation and the other will drift to extinction**
- **The probability of fixation of an allele is equal to its initial frequency**
- **Heterozygosity will decrease over time in a finite population** (it will eventually become 0 when an allele is fixed)
- **Effective population size (N_e)** will determine the effect of genetic drift in a population instead of census size



4

Mutation

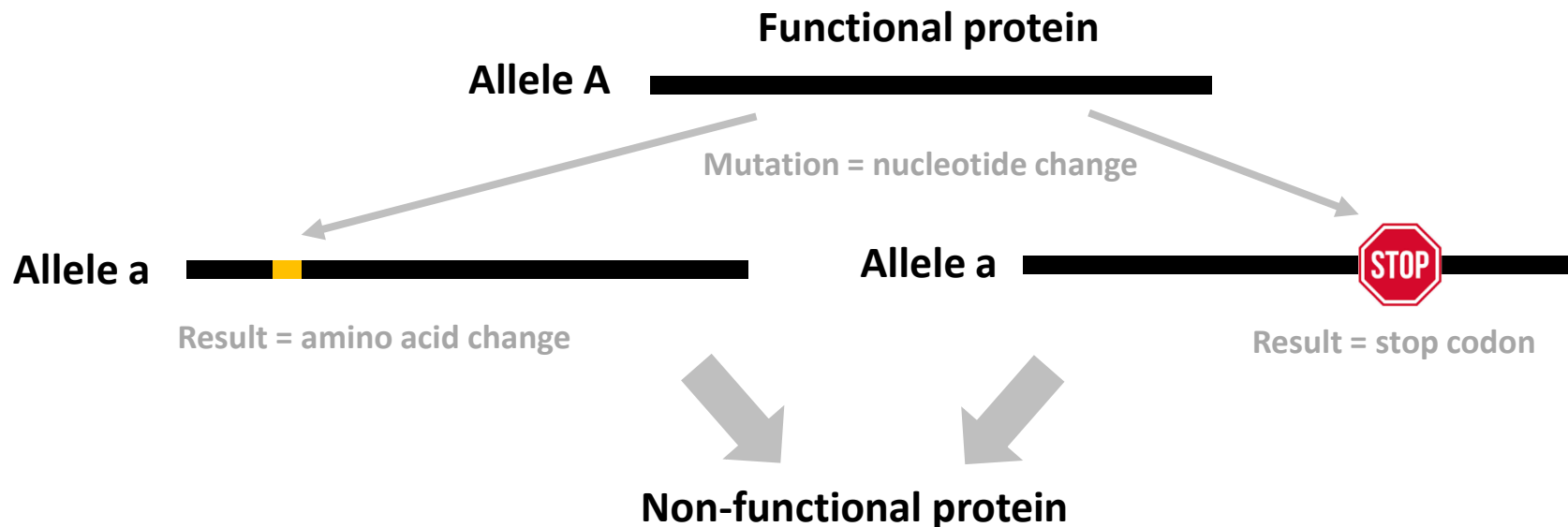
Hardy-Weinberg equilibrium

ASSUMPTIONS

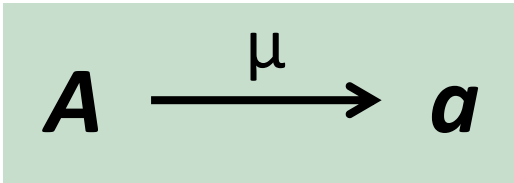
- Diploid organism
- Sexual reproduction
- Non-overlapping generations
- Random mating
- Equal allele frequencies in both sexes
- Large population size
- No migration
- No mutation
- No selection

Mutation as an evolutionary force that changes allele frequencies

- Mutation is the source of all genetic variation. Mutation introduces new alleles in populations.
- A mutation is any permanent change in an organism's DNA (from nucleotide substitutions to large structural variants) and is the result of unrepaired damage in DNA and errors during DNA replication or repair.
- Mutations in the germinal line are transmitted to offspring but somatic mutations are not.
- At phenotypical level, mutation can be considered recurrent. At molecular level most mutations are unique.



Irreversible mutation



$$p_{t+1} = p_t (1 - \mu)$$

$$p_{t+2} = p_{t+1} (1 - \mu)$$

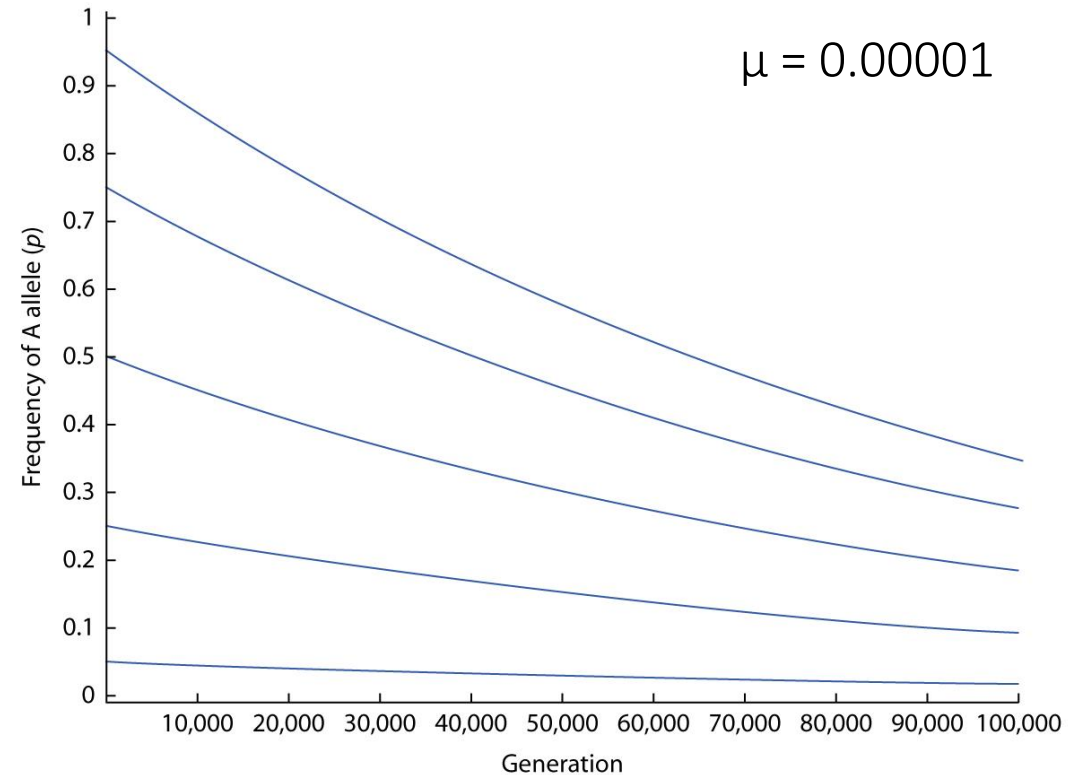
$$p_{t+2} = p_t (1 - \mu) (1 - \mu)$$

$$p_{t+2} = p_t (1 - \mu)^2$$

$$p_{t+3} = p_t (1 - \mu)^3$$

⋮

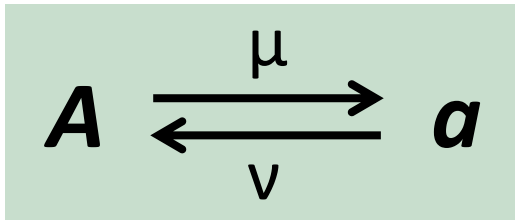
$$p_t = p_0 (1 - \mu)^t$$



When $t \rightarrow \infty$ $p_t \rightarrow 0$ $q_t \rightarrow 1$

Changes in allele frequency due to mutation alone occur over very long time scales

Reversible mutation

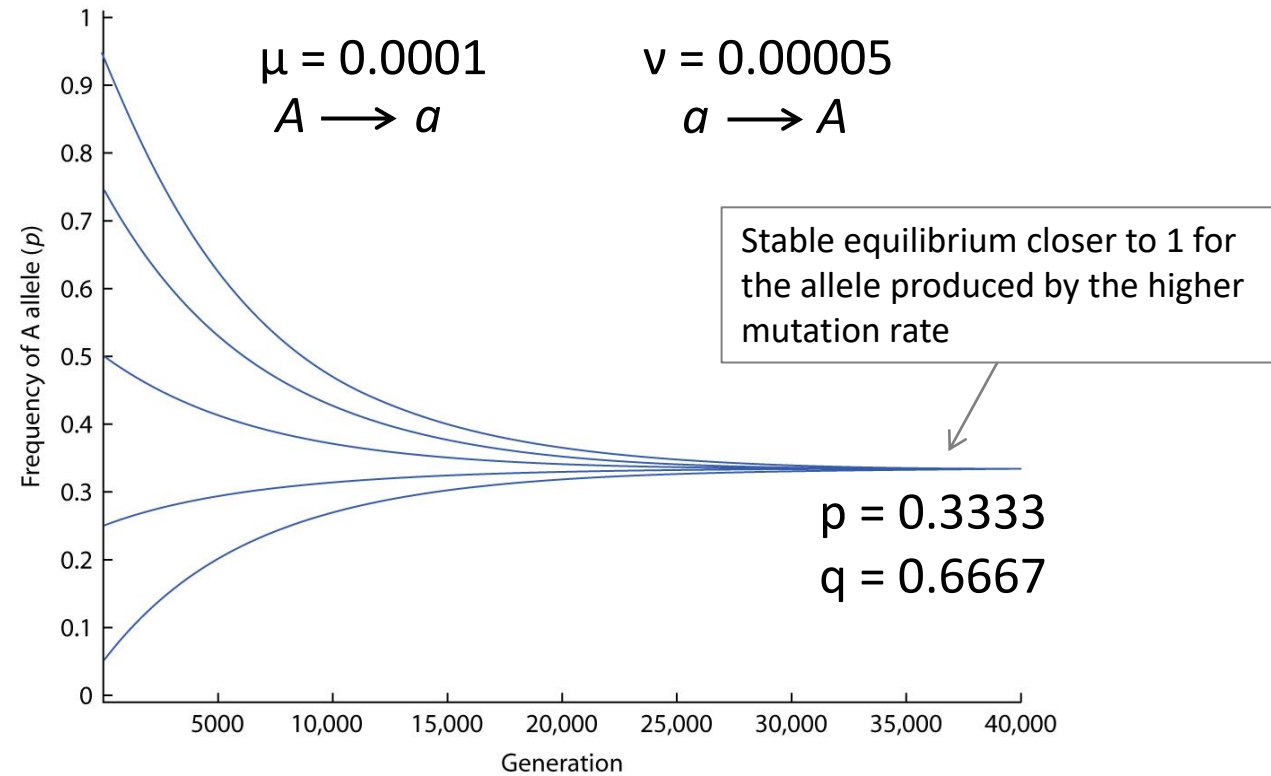


$$p_{t+1} = p_t (1 - \mu) + (1 - p_t) \nu$$

In equilibrium

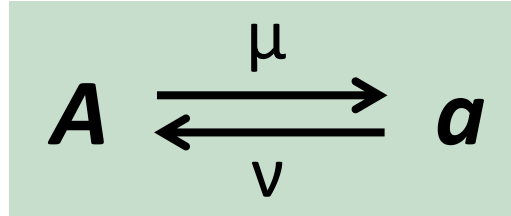
$$\hat{p} = \frac{\nu}{\mu + \nu}$$

$$p_t = \frac{\nu}{\mu + \nu} + \left(p_0 - \frac{\nu}{\mu + \nu} \right) (1 - \mu - \nu)^t$$



When $t \rightarrow \infty$ $(1 - \mu - \nu)^t \rightarrow 0$ $p_t \rightarrow \frac{\nu}{\mu + \nu}$

Reversible mutation



$$\mu > \nu$$

- The rates of mutation from wild type to a novel allele (**forward mutations**) are nearly a factor of 10 more common than mutations from a novel allele to wild type (**reverse mutations**).
- This asymmetry occurs because there are more ways mutation can cause a normal allele to malfunction than there are ways to exactly restore that function once it is disrupted.

