Class project: genetic drift simulation

What is genetic drift?
Why is it important?
How can we model it to understand it better?

Population genetics

- The study of changes in gene frequencies within a population, or between populations, is the field of population genetics
- It is an important topic for:
  - Conservation biology
  - Studies of evolution
  - Plant and animal breeding
  - Biotechnology

Gene basics

- Genes – sequences of DNA, on chromosomes (loci)
  - Some code for proteins
  - Some are regulatory
  - Some (called “pseudogenes”) have no function
- Alleles – alternative forms of a gene
  - Genes will often have more than one allele, sometimes many different ones, in a population
  - An individual can only have at most two different alleles
  - One allele is inherited from mother, one from father
- Homozygote – an individual in which both maternal and paternal copies of a gene are the same allele
- Heterozygote – an individual in which maternal and paternal copies of a gene are different alleles

Importance of genetic diversity

- Genetic diversity is an important part of conserving biological resources
  - Sources of genes for biotechnology
  - Sources of genes for agriculture
  - Sources of genes for disease resistance
- Genetic diversity is also important to the survival of species
  - Individual level = important for immune function (mammalian MHC), reproductive success
  - Population level = increases a population's ability to adapt to changing conditions (new diseases, climate change)

Gene frequencies, allelic diversity

- Gene frequencies = relative number of each allele in a population
- Allelic diversity = number of alleles in the population
- At a population level, a gene is more diverse if:
  - There are more alleles in the population (greater allelic diversity)
  - Allele frequencies are even (for a given allelic diversity)

Alleles for gene A	Copies of allele found in population	Allele frequency
A1	10	10/70 = 0.14
A2	20	20/70 = 0.29
A3	5	5/70 = 0.07
A4	35	35/70 = 0.50
Total	70	

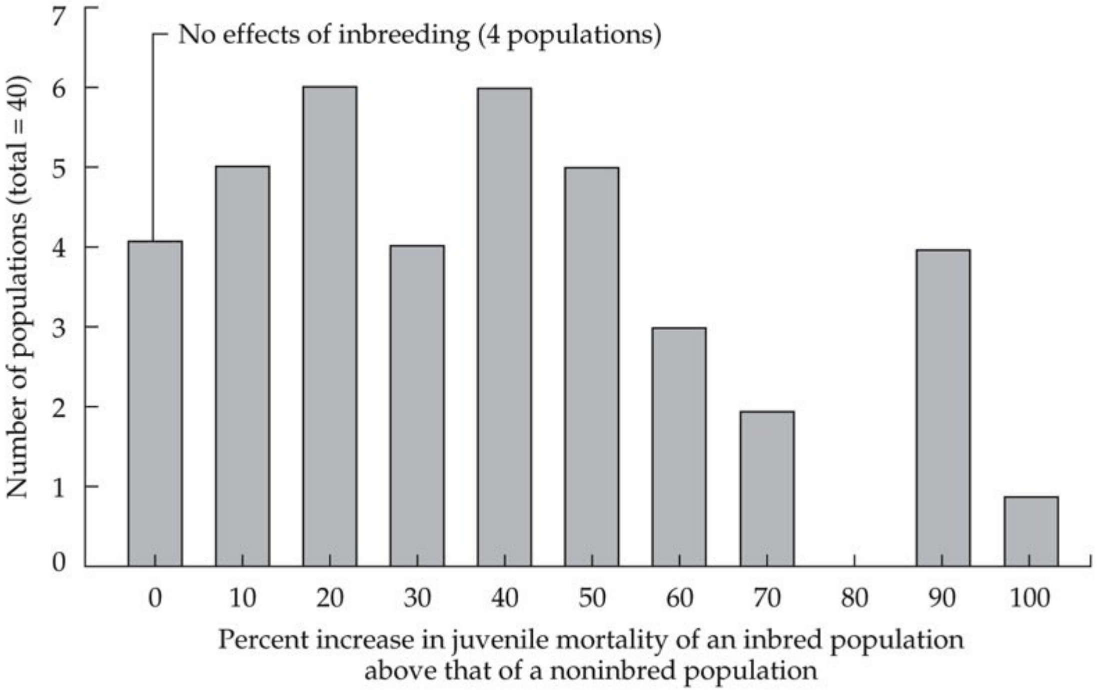
Heterozygosity

- The proportion of individuals that are heterozygous at a locus is the population heterozygosity
- If more than one gene is used, heterozygosity of each gene can be averaged

Loss of heterozygosity can be bad for individuals and populations

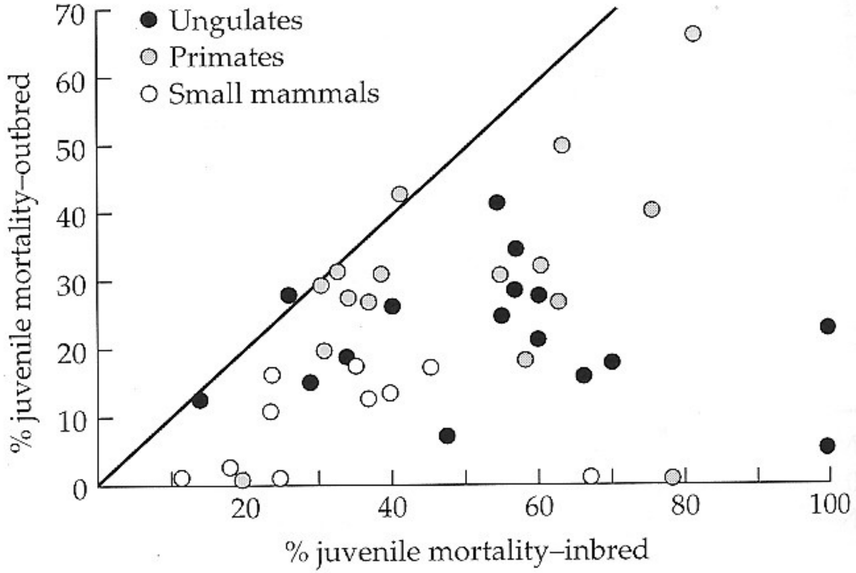
- Most genetic diseases are recessive, high heterozygosity prevents them from being expressed
- Immune function enhanced by high heterozygosity, compromised by low heterozygosity
- Effects include:
  - Decreased fertility
  - High juvenile mortality
  - Sperm, birth defects,
  - Weak, sterile offspring
  - Impaired immune function

Loss of heterozygosity in inbred populations → inbreeding depression



Mammals

ESSENTIALS OF CONSERVATION BIOLOGY, Fourth Edition, Figure 11.6

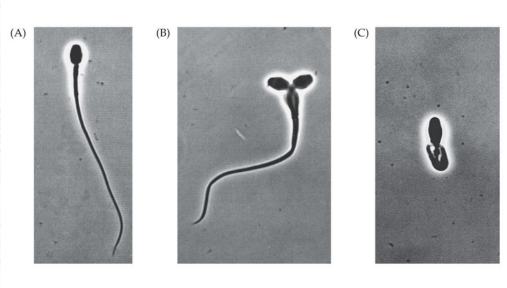


Examples

- Cheetah
  - None of 52 genes tested showed variability - expect 8 to 31%
  - Skin grafts accepted between individuals
  - Abnormal sperm (70%)

- Florida panther (30-50 animals)
  - Sperm defects (93.5%), cryptorchid (testes that don't descend), congenital heart defects, decreased immunity

- Ngorongoro Lion – sperm defects



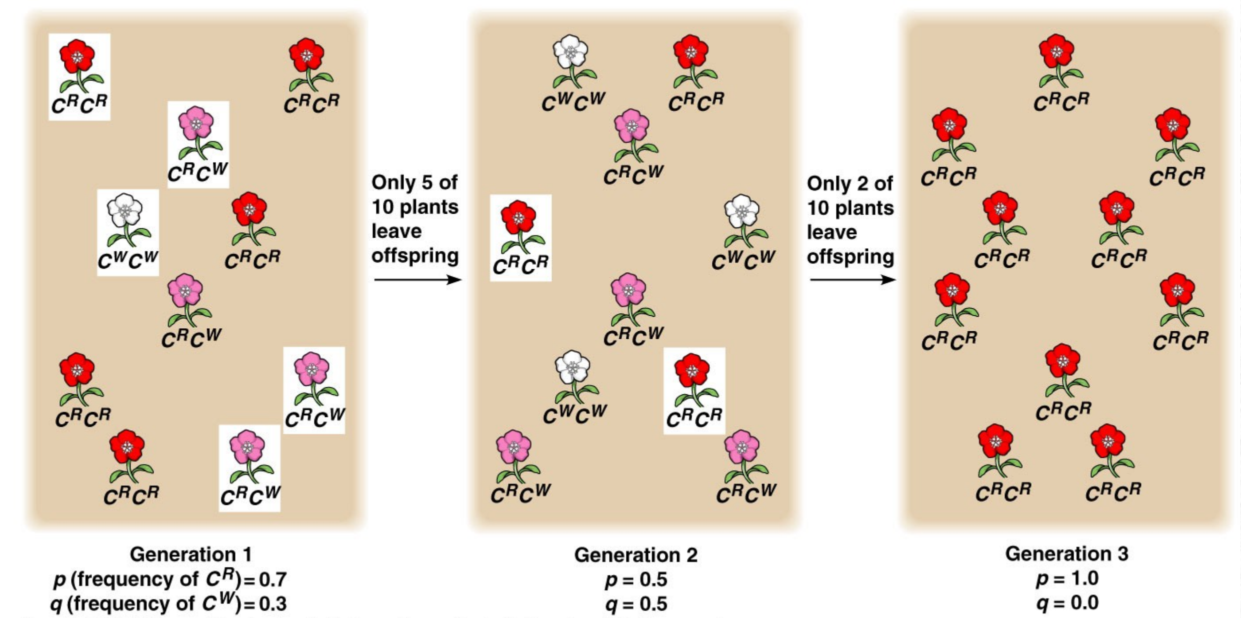
Changes in gene frequencies

- Gene frequencies are a balance between loss and gain of alleles
- Gains in alleles come from:
  - Mutation
  - Immigration (gene flow)
- Losses of alleles come from:
  - **Genetic drift** (bottlenecks, founder effects)
  - Natural selection (under some circumstances)

Changes in allele frequencies due to random mating = genetic drift

- Random mating = choice of mates independent of genotype
- Produces changes in gene frequencies due to random sampling of the available alleles
- Which alleles are passed on, and in what number, is random, unpredictable
- Genetic drift is a form of evolution, but doesn't lead to adaptation

Ball and urn models



http://rosalind.info/media/genetic_drift.gif

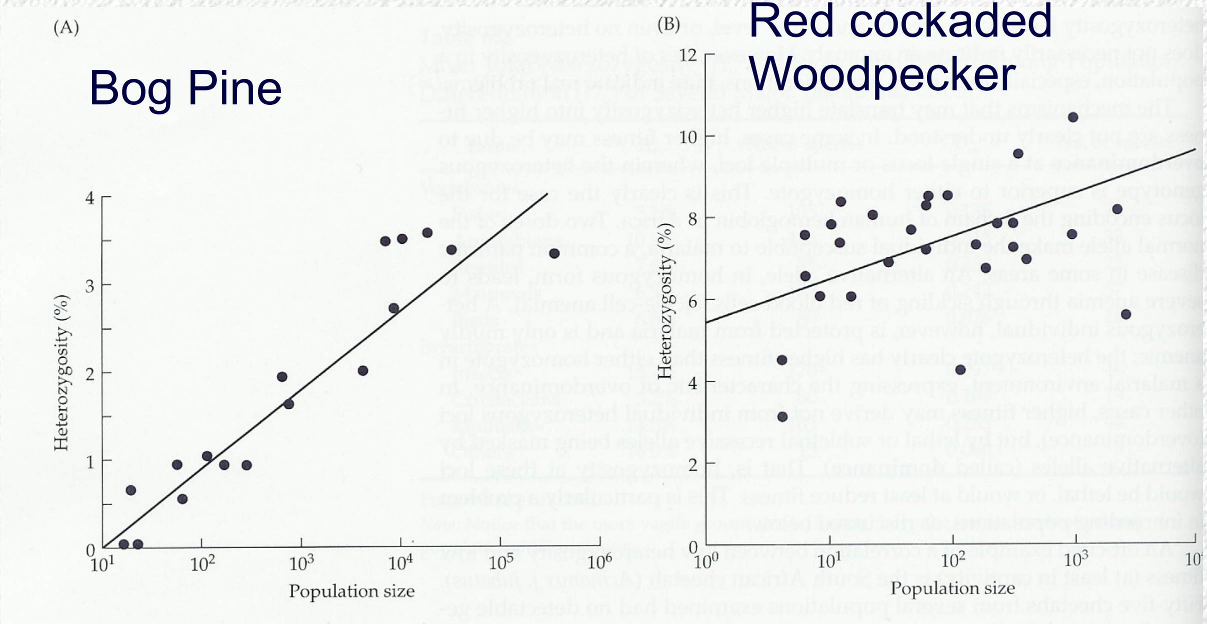
Effect of drift on allelic diversity

- Allelic diversity = number of alleles in the population
- Random changes in allele frequencies can cause loss when:
  - An allele frequency reaches 0 – that allele is lost
  - An allele frequency reaches 1 (**fixation**) – all others are lost
- Frequencies of 0 or 1 are **absorbing states** – once reached, no way to recover in a closed system
  - Can be restored by immigration, mutation
- Fixation of one allele is the expected end result of genetic drift, if it continues long enough, and lost alleles are not restored to the population

How do we study genetic drift?

- Field studies, laboratory studies
  - Look for genetic differences among populations in alleles that don't produce a phenotype (pseudogenes)
  - Or, genes that do produce a phenotype but that don't affect the fitness of the individual
- Mathematical models, computer simulations
  - Study the expected behavior of a gene pool subject to random mating
  - Study effects of various factors on drift (mating systems, population size, fluctuating environments, etc.)

Field studies



Bog Pine

Red cockaded Woodpecker

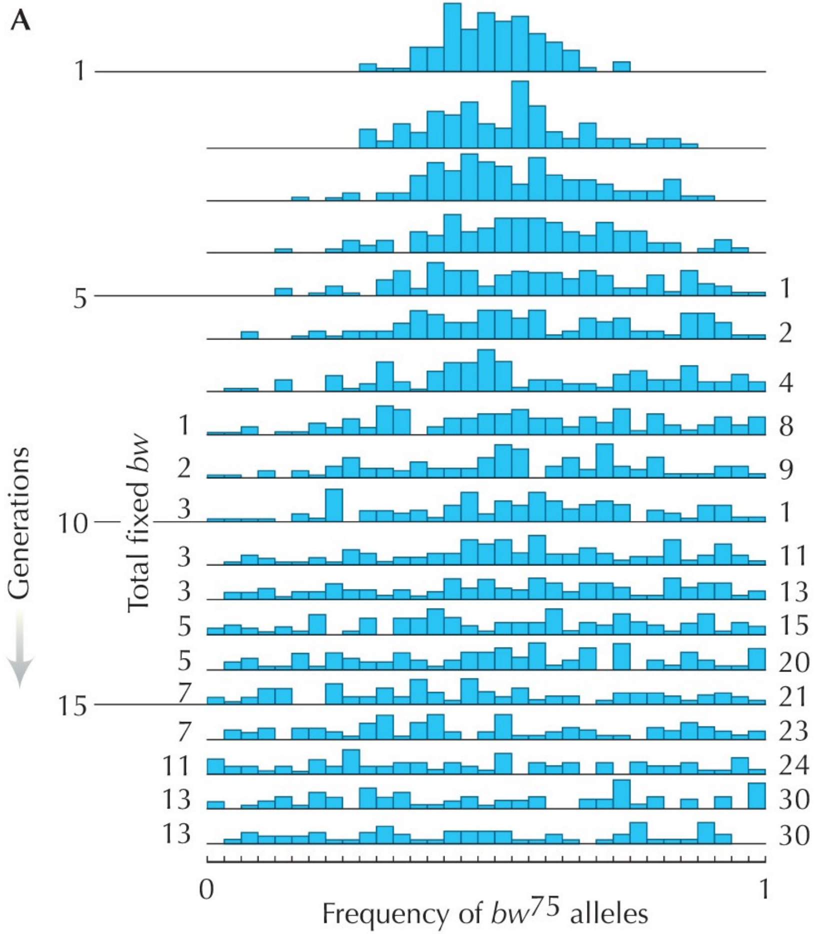


FIGURE 15.3. Random drift in experimental populations of *Drosophila melanogaster*, propagated with eight males and eight females. (*A*) The distribution of allele frequencies across replicate populations, all started at $p = 0.5$. Populations that had fixed one or the other allele are shown at *left* and *right*.

Experiment on lab populations of Drosophila, 8 mating pairs each

Histograms are the gene frequencies for the bw^{75} allele across many populations

All started at a frequency of 0.5

An increasing number of populations became fixed at a frequency of 1 for bw^{75} or 1 for bw, causing loss of the other allele over time

Models of drift

- Deterministic models = equations, no random variation
  - Always the same results given the same inputs
  - Often also analytical = can solve for equilibrium
- Stochastic models = random variation, simulations
  - Same inputs → different outputs due to random chance
  - A single run is unpredictable, but can perform many runs and see what typically happens

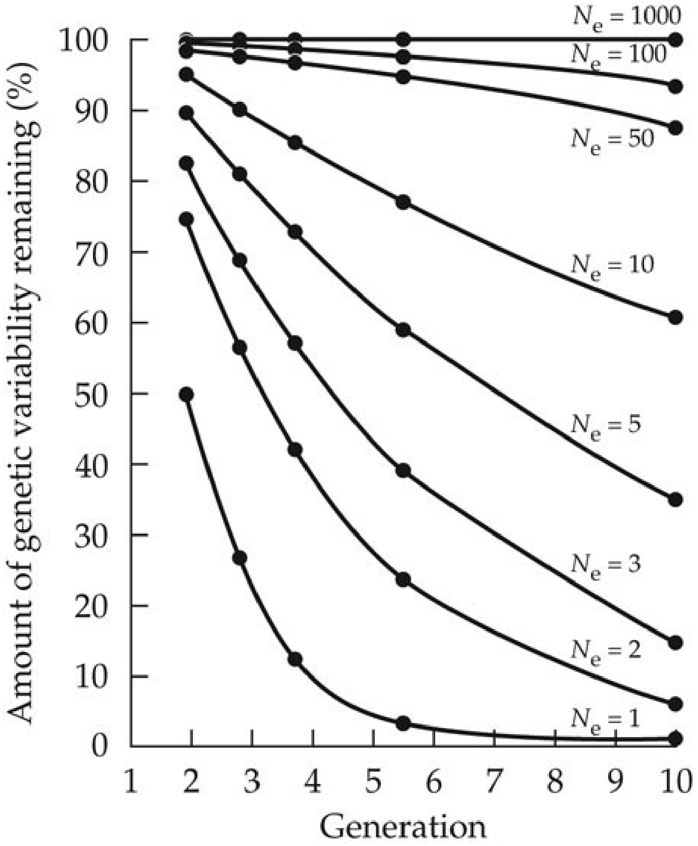
Deterministic model: change in heterozygosity due to drift

- Changes in average heterozygosity over time due to drift – modeled at population level
- Equation: loss of original heterozygosity after each generation due to drift

$$\Delta H = 1 - \frac{1}{2N_e}$$

- Thus, loss in heterozygosity is proportional to $2N_e$
- Deterministic model of a stochastic process – prediction of an average, but there will be variation around it

Proportion HZ remaining



ESSENTIALS OF CONSERVATION BIOLOGY, Fourth Edition, Figure 11.3 © 2006 Sinauer Associates, Inc.

Deterministic models don’t capture the uncertainty in the outcome of drift

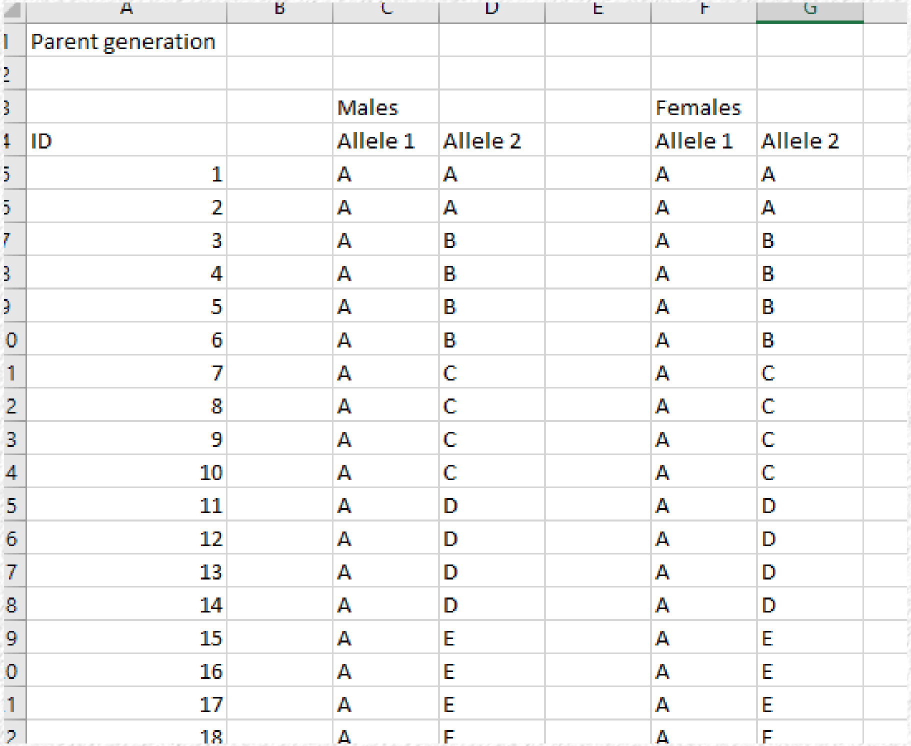
- Drift is a **stochastic** process
- The expected (average) outcome is only the most likely one, but not the only possibility
- Actual rates change in the gene pool can be much more rapid, or much slower
- To capture this unpredictability, it is useful to use a stochastic simulation model
  - Multiple runs
  - See what the range of possible outcomes looks like, as well as the expected outcome

Our stochastic simulation model of drift

- Our model will be:
  - **Individual-based**, or **agent-based** – each individual in the population will be simulated
  - Stochastic – not deterministic, mates will be done at random
  - Simulation – not analytical, results will be studied
- We will assess the effects on drift of:
  - Population size
  - Immigration

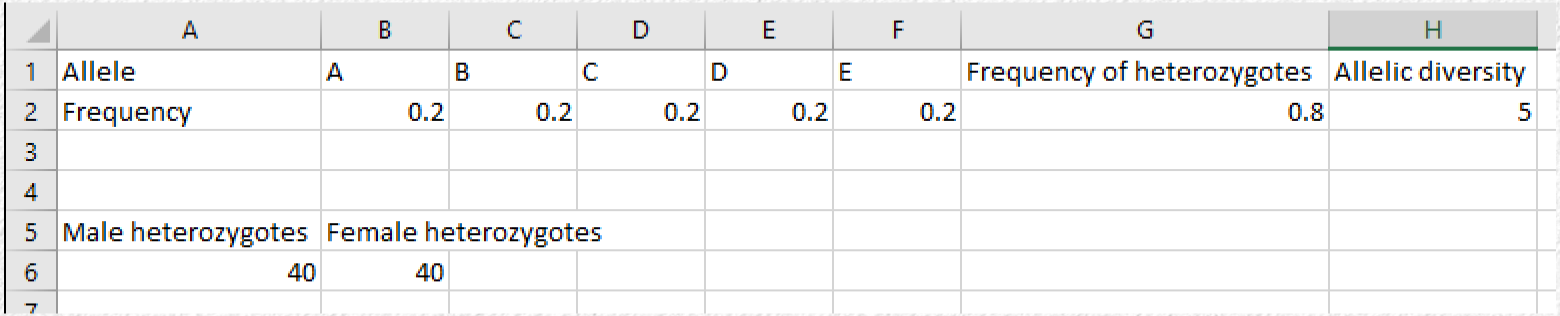
Steps

- Set up an initial population with equal frequencies of 5 alleles
  - 50 males, 50 females



A	B	C	D	E	F	G
Parent generation						
		Males			Females	
ID		Allele 1	Allele 2		Allele 1	Allele 2
1		A	A		A	A
2		A	A		A	A
3		A	B		A	B
4		A	B		A	B
5		A	B		A	B
6		A	B		A	B
7		A	C		A	C
8		A	C		A	C
9		A	C		A	C
10		A	C		A	C
11		A	D		A	D
12		A	D		A	D
13		A	D		A	D
14		A	D		A	D
15		A	E		A	E
16		A	E		A	E
17		A	E		A	E
18		A	E		A	E

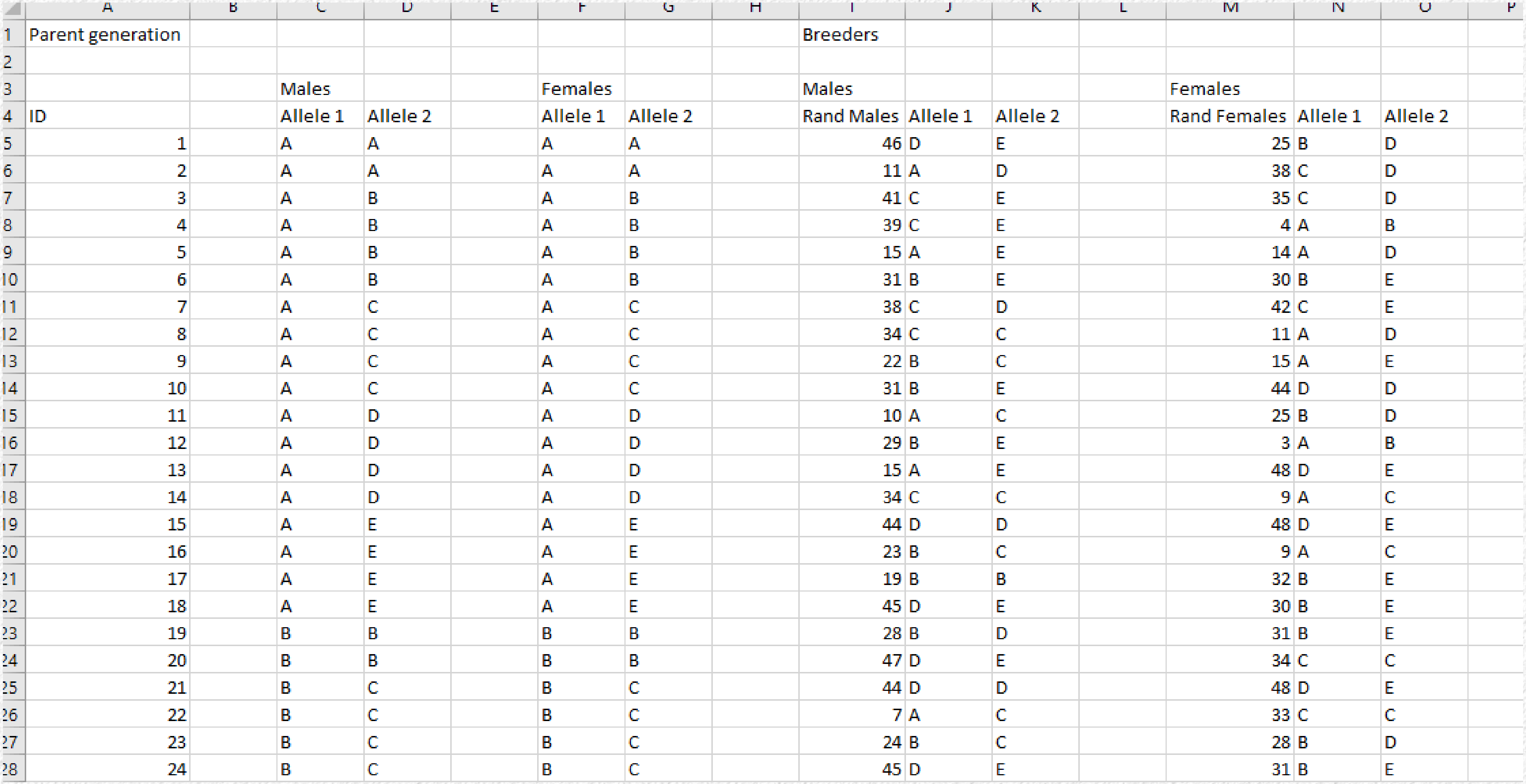
Based on Hardy-Weinberg frequencies



	A	B	C	D	E	F	G	H
1	Allele	A	B	C	D	E	Frequency of heterozygotes	Allelic diversity
2	Frequency	0.2	0.2	0.2	0.2	0.2	0.8	5
3								
4								
5	Male heterozygotes	Female heterozygotes						
6	40	40						

Steps

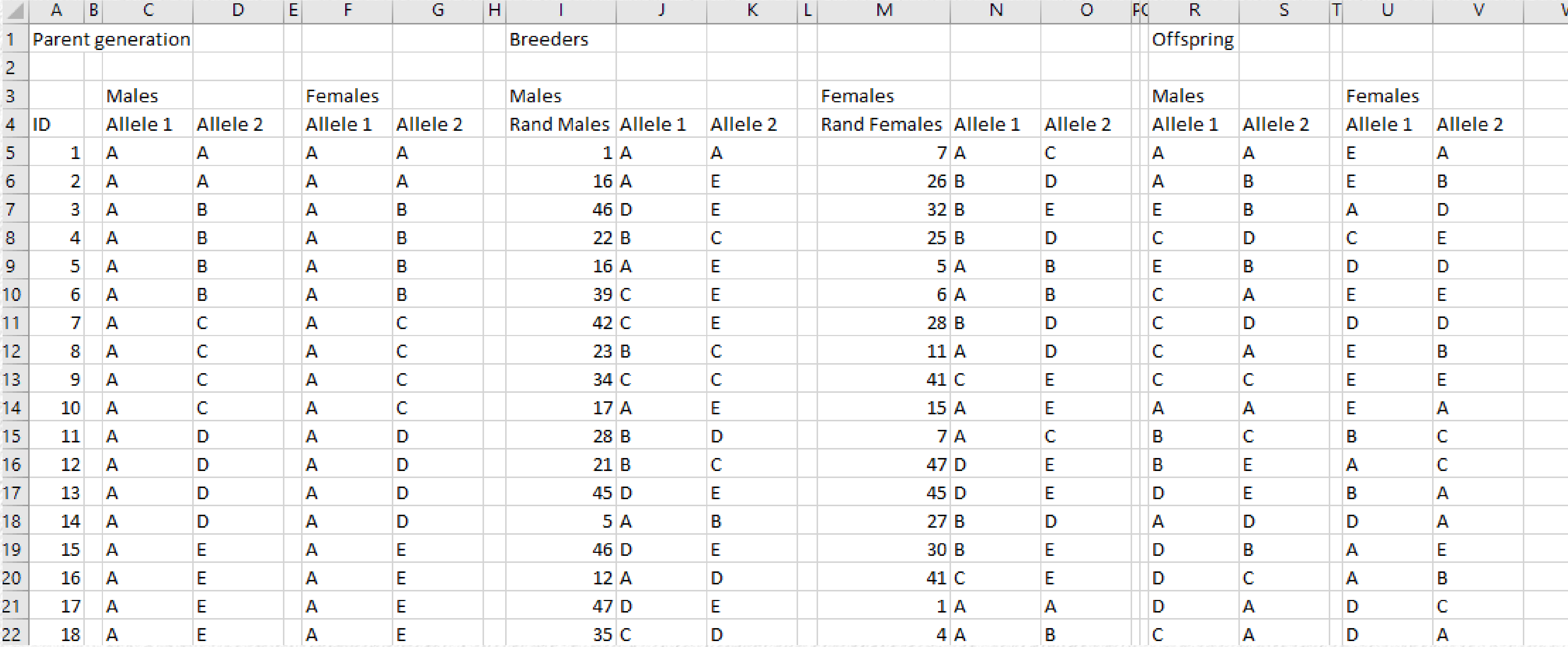
- Set up an initial population with equal frequencies of 5 alleles
  - 50 males, 50 females
- For 500 generations:
  - Randomly select breeders from the current population



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Parent generation								Breeders							
2																
3			Males			Females			Males				Females			
4	ID		Allele 1	Allele 2		Allele 1	Allele 2		Rand Males	Allele 1	Allele 2		Rand Females	Allele 1	Allele 2	
5	1		A	A		A	A		46	D	E		25	B	D	
6	2		A	A		A	A		11	A	D		38	C	D	
7	3		A	B		A	B		41	C	E		35	C	D	
8	4		A	B		A	B		39	C	E		4	A	B	
9	5		A	B		A	B		15	A	E		14	A	D	
10	6		A	B		A	B		31	B	E		30	B	E	
11	7		A	C		A	C		38	C	D		42	C	E	
12	8		A	C		A	C		34	C	C		11	A	D	
13	9		A	C		A	C		22	B	C		15	A	E	
14	10		A	C		A	C		31	B	E		44	D	D	
15	11		A	D		A	D		10	A	C		25	B	D	
16	12		A	D		A	D		29	B	E		3	A	B	
17	13		A	D		A	D		15	A	E		48	D	E	
18	14		A	D		A	D		34	C	C		9	A	C	
19	15		A	E		A	E		44	D	D		48	D	E	
20	16		A	E		A	E		23	B	C		9	A	C	
21	17		A	E		A	E		19	B	B		32	B	E	
22	18		A	E		A	E		45	D	E		30	B	E	
23	19		B	B		B	B		28	B	D		31	B	E	
24	20		B	B		B	B		47	D	E		34	C	C	
25	21		B	C		B	C		44	D	D		48	D	E	
26	22		B	C		B	C		7	A	C		33	C	C	
27	23		B	C		B	C		24	B	C		28	B	D	
28	24		B	C		B	C		45	D	E		31	B	E	

Steps

- Set up an initial population with equal frequencies of 5 alleles
  - 50 males, 50 females
- For 500 generations:
  - Randomly select breeders from the current population
  - Mate breeders to produce 50 female offspring
  - Select new breeders, mate them to produce 50 male offspring

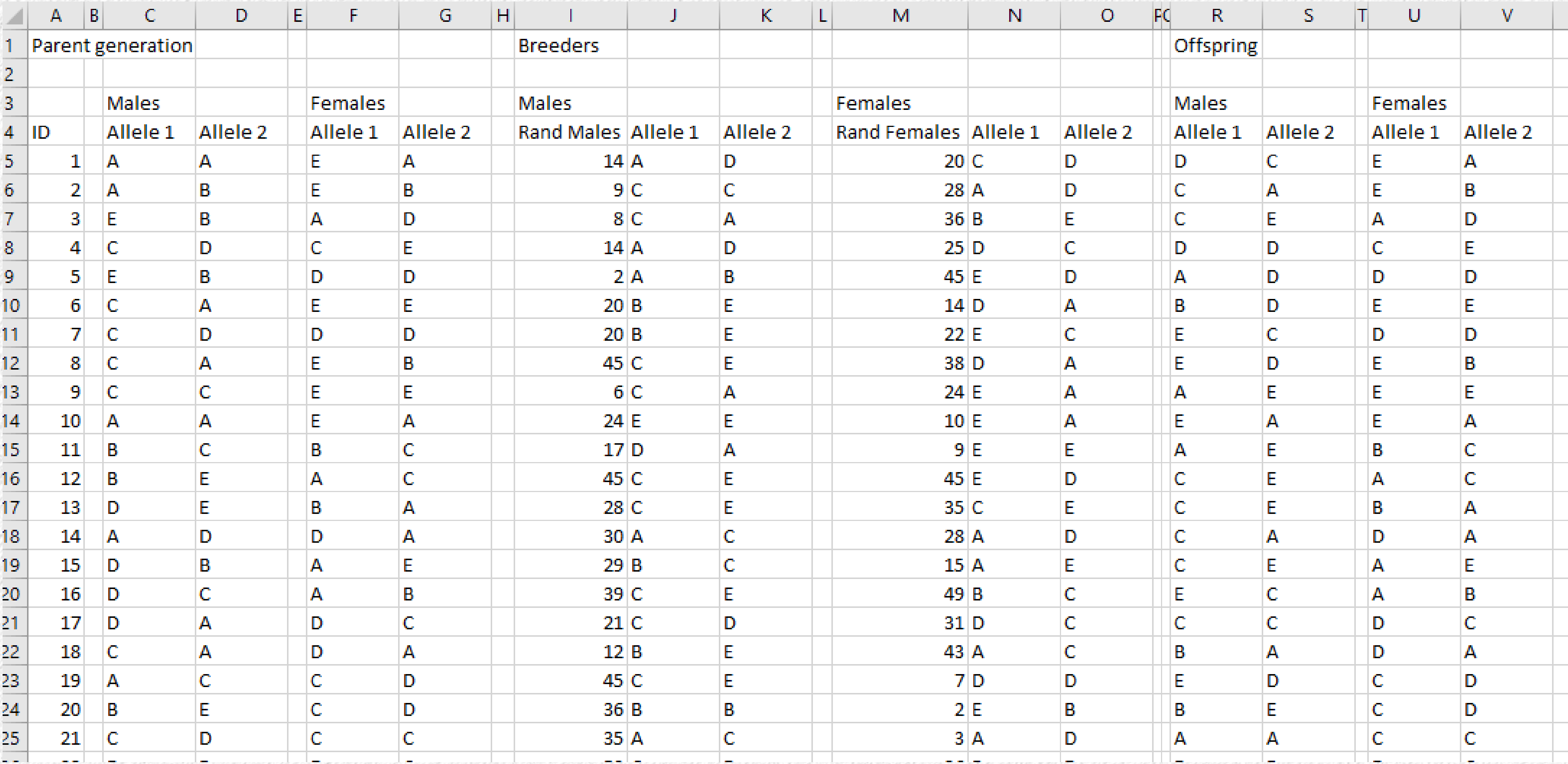


	A	C	D	F	G	I	J	K	M	N	O	R	S	U	V
1	Parent generation					Breeders						Offspring			
2															
3		Males		Females		Males			Females			Males		Females	
4	ID	Allele 1	Allele 2	Allele 1	Allele 2	Rand Males	Allele 1	Allele 2	Rand Females	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2
5	1	A	A	A	A	1	A	A	7	A	C	A	A	E	A
6	2	A	A	A	A	16	A	E	26	B	D	A	B	E	B
7	3	A	B	A	B	46	D	E	32	B	E	E	B	A	D
8	4	A	B	A	B	22	B	C	25	B	D	C	D	C	E
9	5	A	B	A	B	16	A	E	5	A	B	E	B	D	D
10	6	A	B	A	B	39	C	E	6	A	B	C	A	E	E
11	7	A	C	A	C	42	C	E	28	B	D	C	D	D	D
12	8	A	C	A	C	23	B	C	11	A	D	C	A	E	B
13	9	A	C	A	C	34	C	C	41	C	E	C	C	E	E
14	10	A	C	A	C	17	A	E	15	A	E	A	A	E	A
15	11	A	D	A	D	28	B	D	7	A	C	B	C	B	C
16	12	A	D	A	D	21	B	C	47	D	E	B	E	A	C
17	13	A	D	A	D	45	D	E	45	D	E	D	E	B	A
18	14	A	D	A	D	5	A	B	27	B	D	A	D	D	A
19	15	A	E	A	E	46	D	E	30	B	E	D	B	A	E
20	16	A	E	A	E	12	A	D	41	C	E	D	C	A	B
21	17	A	E	A	E	47	D	E	1	A	A	D	A	D	C
22	18	A	E	A	E	35	C	D	4	A	B	C	A	D	A

Offspring genotypes are randomly selected from parents'

Steps

- Set up an initial population with equal frequencies of 5 alleles
  - 50 males, 50 females
- For 500 generations, do:
  - Randomly select breeders from the current population
  - Mate breeders to produce 50 female offspring
  - Select new breeders, mate them to produce 50 male offspring
  - Offspring become next generation’s adults

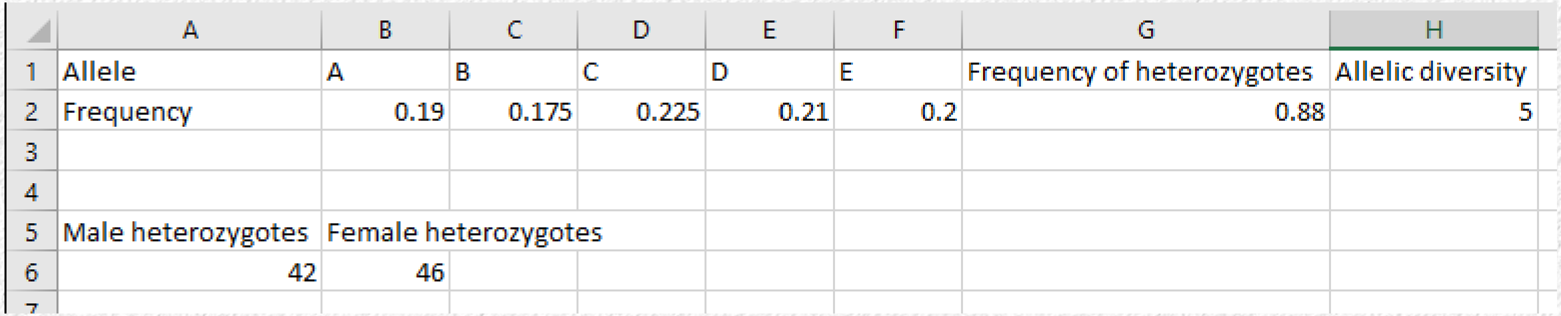


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	PQ	R	S	T	U	V
1	Parent generation								Breeders								Offspring				
2																					
3			Males			Females			Males				Females				Males			Females	
4	ID		Allele 1	Allele 2		Allele 1	Allele 2		Rand Males	Allele 1	Allele 2		Rand Females	Allele 1	Allele 2		Allele 1	Allele 2		Allele 1	Allele 2
5	1		A	A		E	A		14	A	D		20	C	D		D	C		E	A
6	2		A	B		E	B		9	C	C		28	A	D		C	A		E	B
7	3		E	B		A	D		8	C	A		36	B	E		C	E		A	D
8	4		C	D		C	E		14	A	D		25	D	C		D	D		C	E
9	5		E	B		D	D		2	A	B		45	E	D		A	D		D	D
10	6		C	A		E	E		20	B	E		14	D	A		B	D		E	E
11	7		C	D		D	D		20	B	E		22	E	C		E	C		D	D
12	8		C	A		E	B		45	C	E		38	D	A		E	D		E	B
13	9		C	C		E	E		6	C	A		24	E	A		A	E		E	E
14	10		A	A		E	A		24	E	E		10	E	A		E	A		E	A
15	11		B	C		B	C		17	D	A		9	E	E		A	E		B	C
16	12		B	E		A	C		45	C	E		45	E	D		C	E		A	C
17	13		D	E		B	A		28	C	E		35	C	E		C	E		B	A
18	14		A	D		D	A		30	A	C		28	A	D		C	A		D	A
19	15		D	B		A	E		29	B	C		15	A	E		C	E		A	E
20	16		D	C		A	B		39	C	E		49	B	C		E	C		A	B
21	17		D	A		D	C		21	C	D		31	D	C		C	C		D	C
22	18		C	A		D	A		12	B	E		43	A	C		B	A		D	A
23	19		A	C		C	D		45	C	E		7	D	D		E	D		C	D
24	20		B	E		C	D		36	B	B		2	E	B		B	E		C	D
25	21		C	D		C	C		35	A	C		3	A	D		A	A		C	C

Offspring offspring copied/pasted to parent generation

Steps

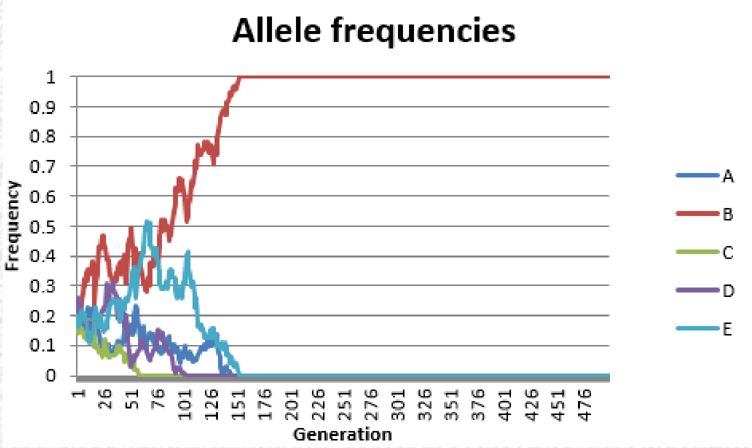
- Set up an initial population with equal frequencies of 5 alleles
  - 50 males, 50 females
- For 500 generations, do:
  - Randomly select breeders from the current population
  - Mate breeders to produce 50 female offspring
  - Select new breeders, mate them to produce 50 male offspring
  - Offspring become next generation’s adults
  - Measure the allele frequencies, heterozygosity

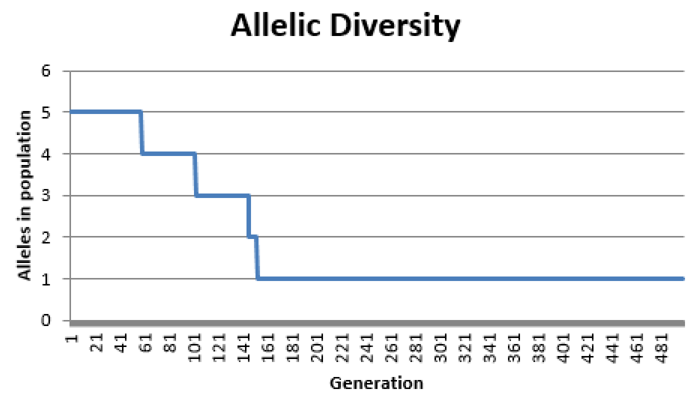


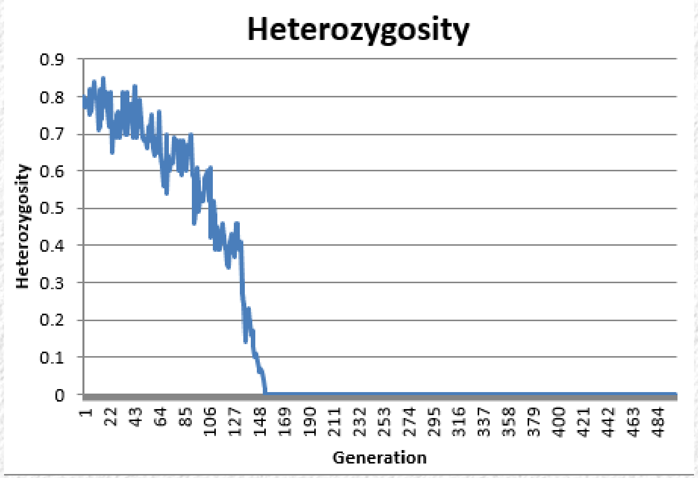
	A	B	C	D	E	F	G	H
1	Allele	A	B	C	D	E	Frequency of heterozygotes	Allelic diversity
2	Frequency	0.19	0.175	0.225	0.21	0.2	0.88	5
3								
4								
5	Male heterozygotes	Female heterozygotes						
6	42	46						

Change in frequencies, heterozygosity, and number of alleles after one generation

After 500 generations:







Change in frequencies, heterozygosity, and number of alleles after 500 generations

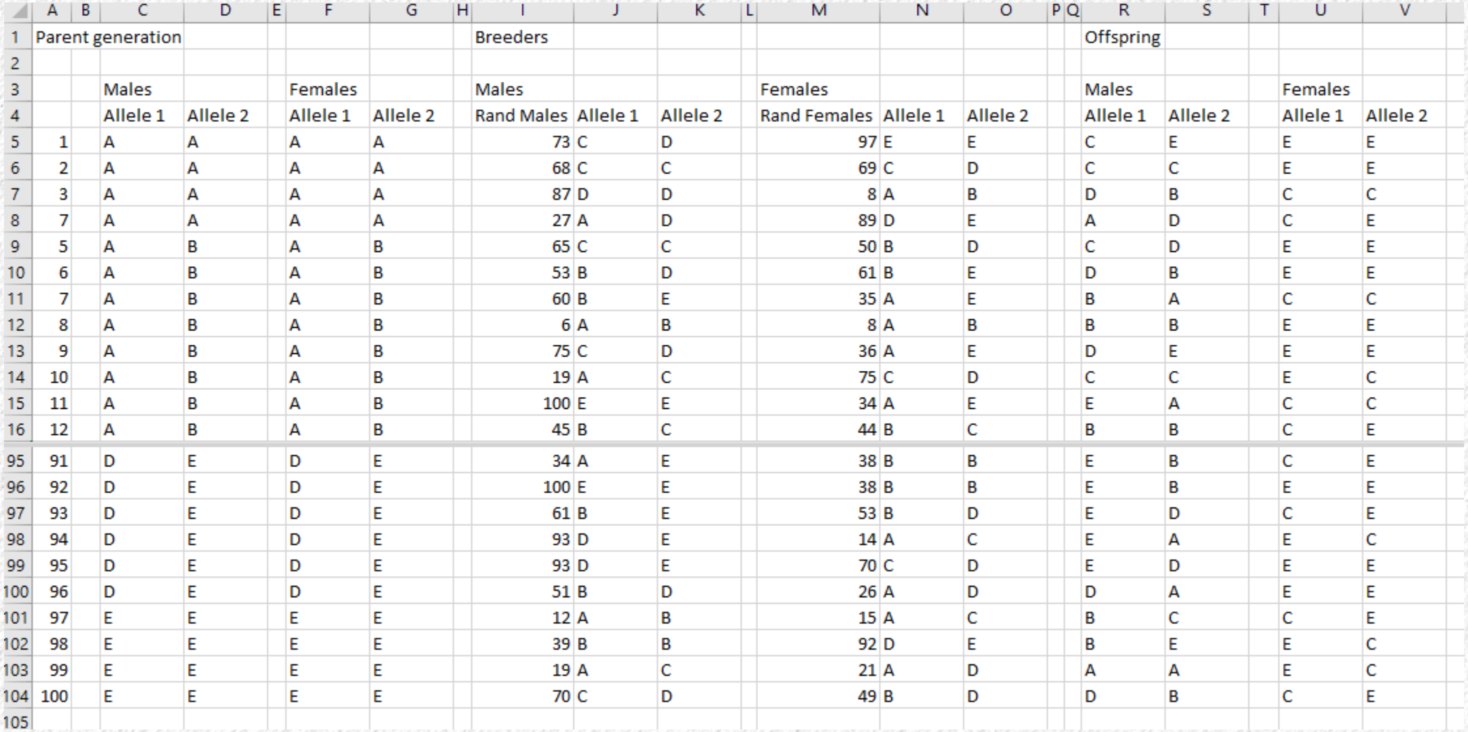
But, this is a stochastic simulation, results won’t always be the same

- For example, this
- What to do? Run the simulation 100 times, record:
  - Which allele goes to fixation
  - The generation it happens
- Across the 100 runs summarize:
  - Mean amount of time to fixation
  - How often each allele goes to fixation

Effects of population size, immigration

- Population size:
  - Double the population to 100 males and 100 females and re-run the simulation
- Immigration:
  - Allow 1 male immigrant per generation as a breeder, re-run the simulation

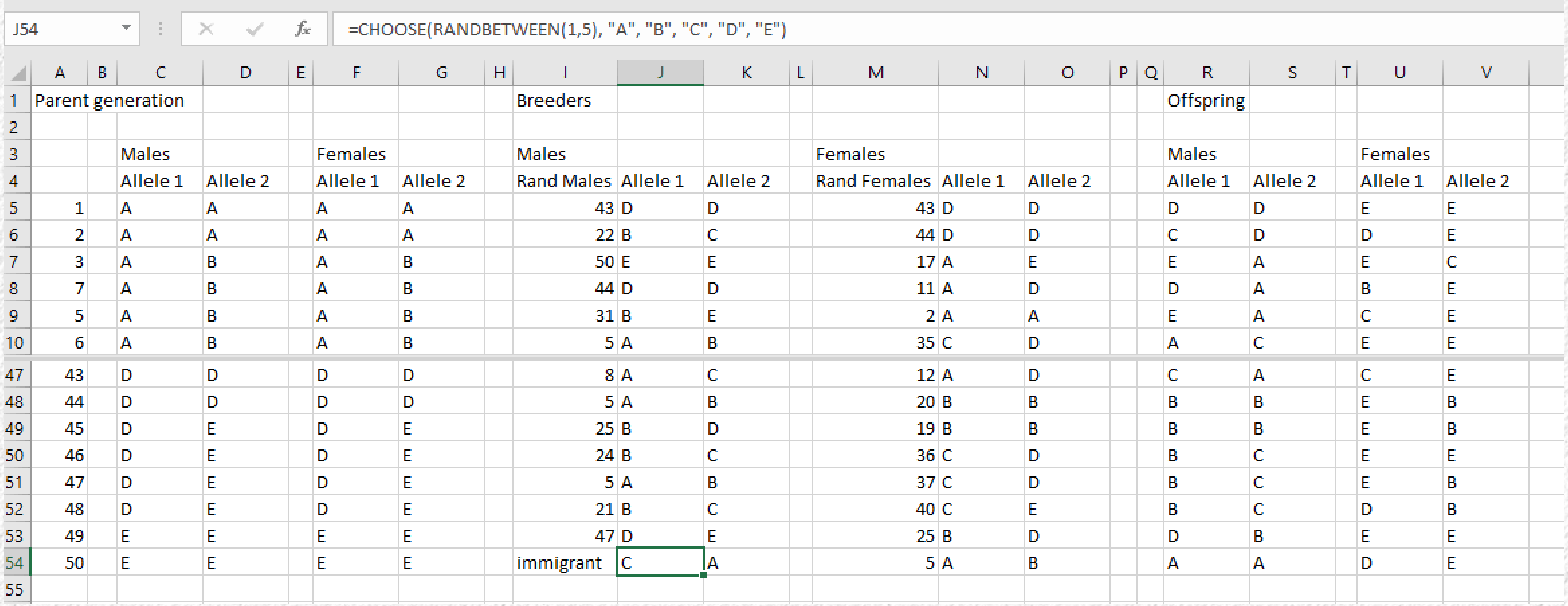
Bigger population size



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Parent generation								Breeders									Offspring				
2																						
3			Males			Females			Males				Females					Males			Females	
4			Allele 1	Allele 2		Allele 1	Allele 2		Rand Males	Allele 1	Allele 2		Rand Females	Allele 1	Allele 2			Allele 1	Allele 2		Allele 1	Allele 2
5	1		A	A		A	A		73	C	D		97	E	E			C	E		E	E
6	2		A	A		A	A		68	C	C		69	C	D			C	C		E	E
7	3		A	A		A	A		87	D	D		8	A	B			D	B		C	C
8	7		A	A		A	A		27	A	D		89	D	E			A	D		C	E
9	5		A	B		A	B		65	C	C		50	B	D			C	D		E	E
10	6		A	B		A	B		53	B	D		61	B	E			D	B		E	E
11	7		A	B		A	B		60	B	E		35	A	E			B	A		C	C
12	8		A	B		A	B		6	A	B		8	A	B			B	B		E	E
13	9		A	B		A	B		75	C	D		36	A	E			D	E		E	E
14	10		A	B		A	B		19	A	C		75	C	D			C	C		E	C
15	11		A	B		A	B		100	E	E		34	A	E			E	A		C	C
16	12		A	B		A	B		45	B	C		44	B	C			B	B		C	E
95	91		D	E		D	E		34	A	E		38	B	B			E	B		C	E
96	92		D	E		D	E		100	E	E		38	B	B			E	B		E	E
97	93		D	E		D	E		61	B	E		53	B	D			E	D		C	E
98	94		D	E		D	E		93	D	E		14	A	C			E	A		E	C
99	95		D	E		D	E		93	D	E		70	C	D			E	D		E	E
100	96		D	E		D	E		51	B	D		26	A	D			D	A		E	E
101	97		E	E		E	E		12	A	B		15	A	C			B	C		C	E
102	98		E	E		E	E		39	B	B		92	D	E			B	E		E	C
103	99		E	E		E	E		19	A	C		21	A	D			A	A		E	C
104	100		E	E		E	E		70	C	D		49	B	D			D	B		C	E
105																						

Same setup, but with 100 males and 100 females – what should happen?

Immigration



J54

=CHOOSE(RANDBETWEEN(1,5), "A", "B", "C", "D", "E")

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Parent generation								Breeders									Offspring				
2																						
3			Males			Females			Males				Females					Males			Females	
4			Allele 1	Allele 2		Allele 1	Allele 2		Rand Males	Allele 1	Allele 2		Rand Females	Allele 1	Allele 2			Allele 1	Allele 2		Allele 1	Allele 2
5	1		A	A		A	A		43	D	D		43	D	D			D	D		E	E
6	2		A	A		A	A		22	B	C		44	D	D			C	D		D	E
7	3		A	B		A	B		50	E	E		17	A	E			E	A		E	C
8	7		A	B		A	B		44	D	D		11	A	D			D	A		B	E
9	5		A	B		A	B		31	B	E		2	A	A			E	A		C	E
10	6		A	B		A	B		5	A	B		35	C	D			A	C		E	E
47	43		D	D		D	D		8	A	C		12	A	D			C	A		C	E
48	44		D	D		D	D		5	A	B		20	B	B			B	B		E	B
49	45		D	E		D	E		25	B	D		19	B	B			B	B		E	B
50	46		D	E		D	E		24	B	C		36	C	D			B	C		E	E
51	47		D	E		D	E		5	A	B		37	C	D			B	C		E	B
52	48		D	E		D	E		21	B	C		40	C	E			B	C		D	B
53	49		E	E		E	E		47	D	E		25	B	D			D	B		E	E
54	50		E	E		E	E		immigrant	C	A		5	A	B			A	A		D	E
55																						

50 males, 50 females, but one male breeder each generation is sampled from a population with equal frequencies for all alleles

What should happen? Are frequencies of 0 and 1 still absorbing states?