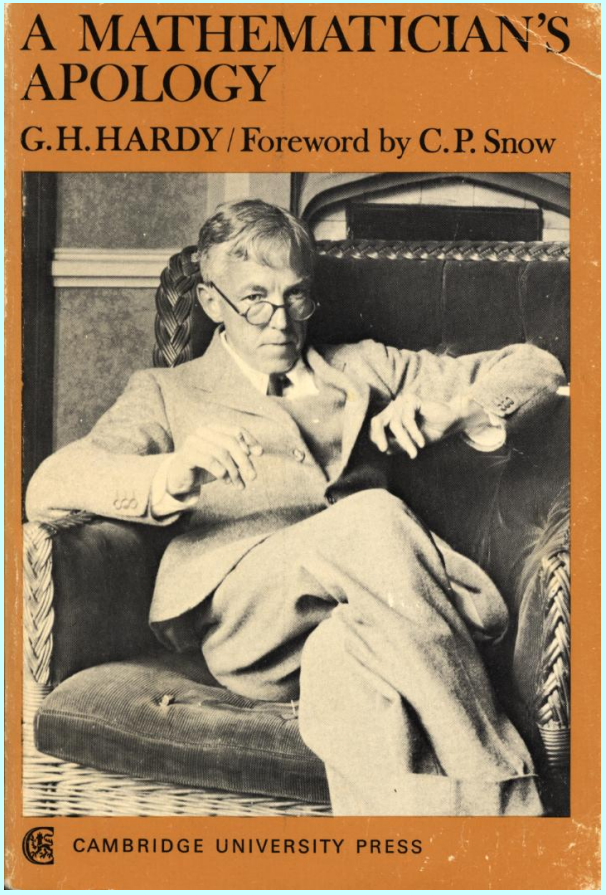
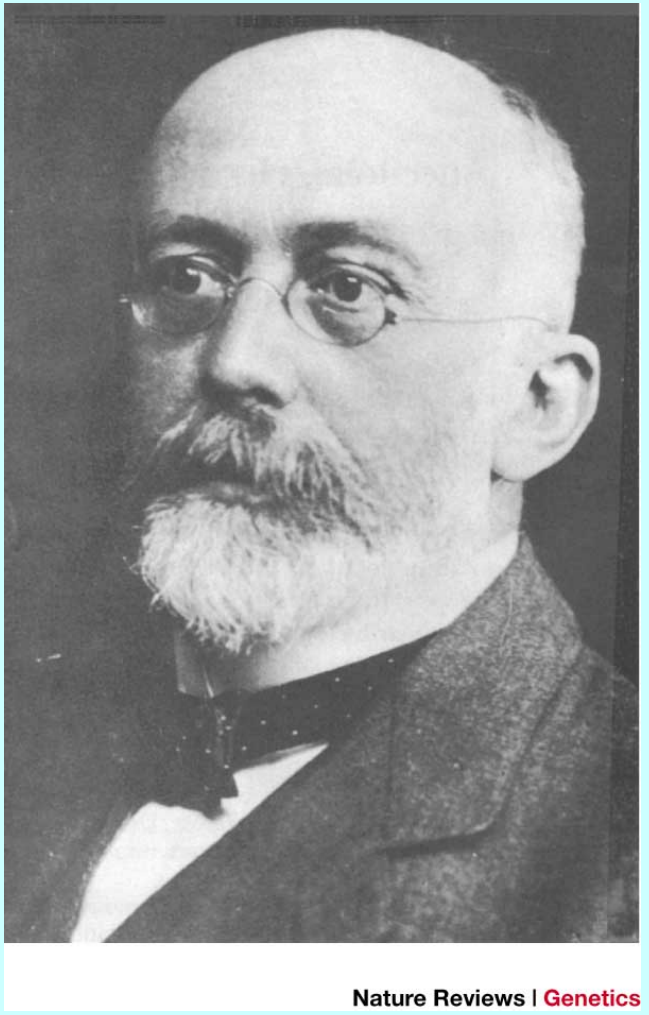
Population Genetics

Joe Felsenstein

GENOME 453, Autumn 2013



Godfrey Harold Hardy (1877-1947)

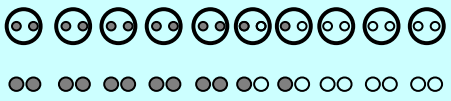


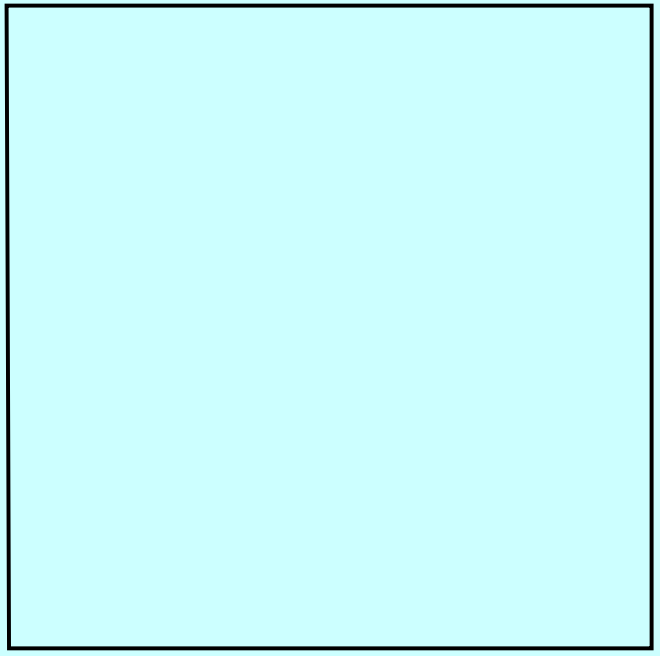
Wilhelm Weinberg (1862-1937)

A Hardy-Weinberg calculation

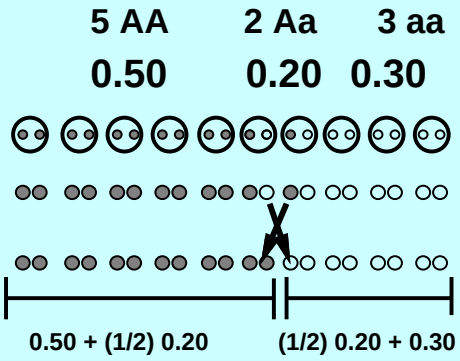
5 AA **2 Aa** **3 aa**

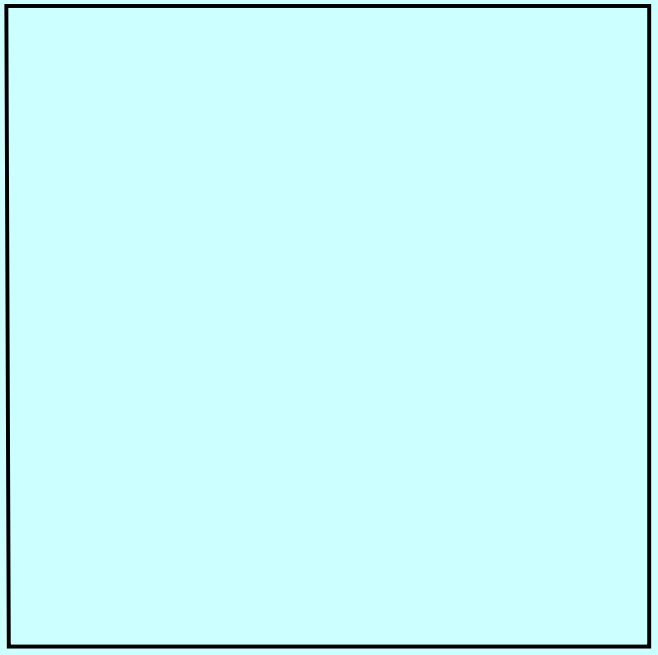
0.50 **0.20** **0.30**



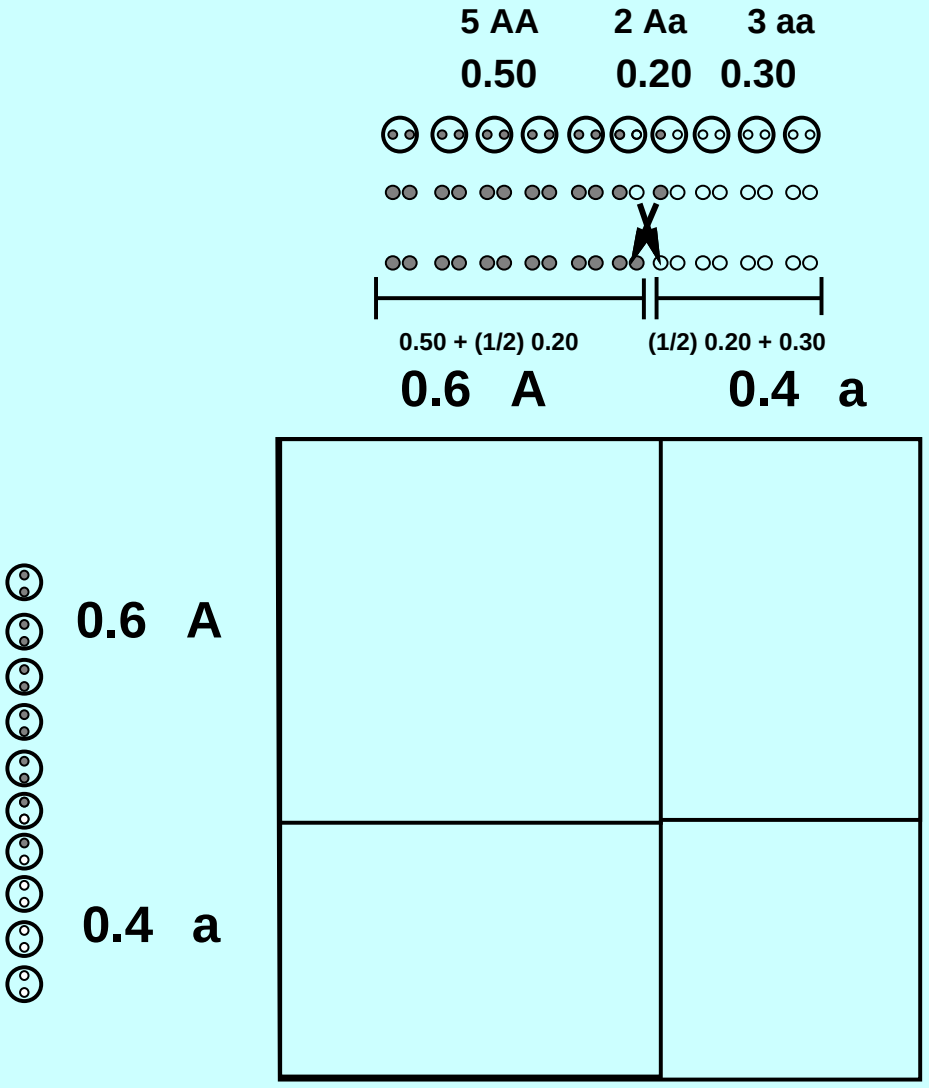


A Hardy-Weinberg calculation

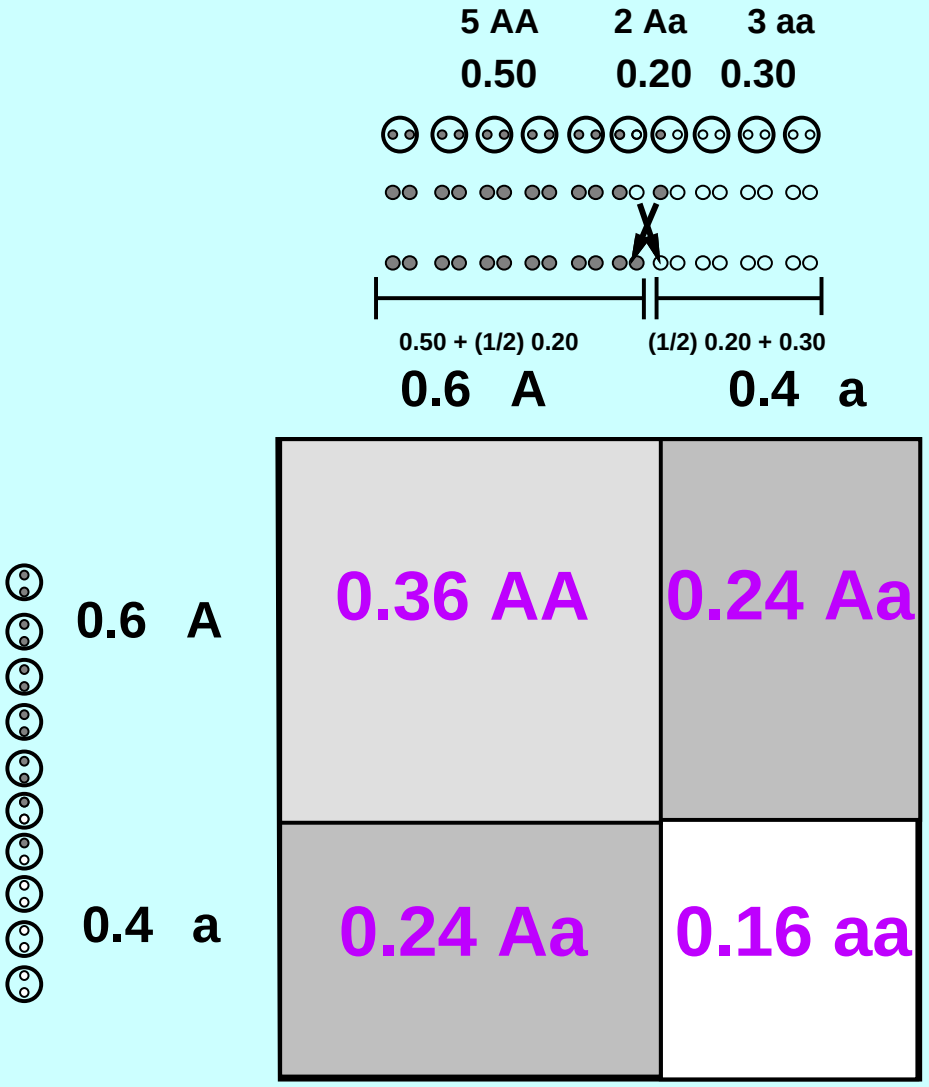




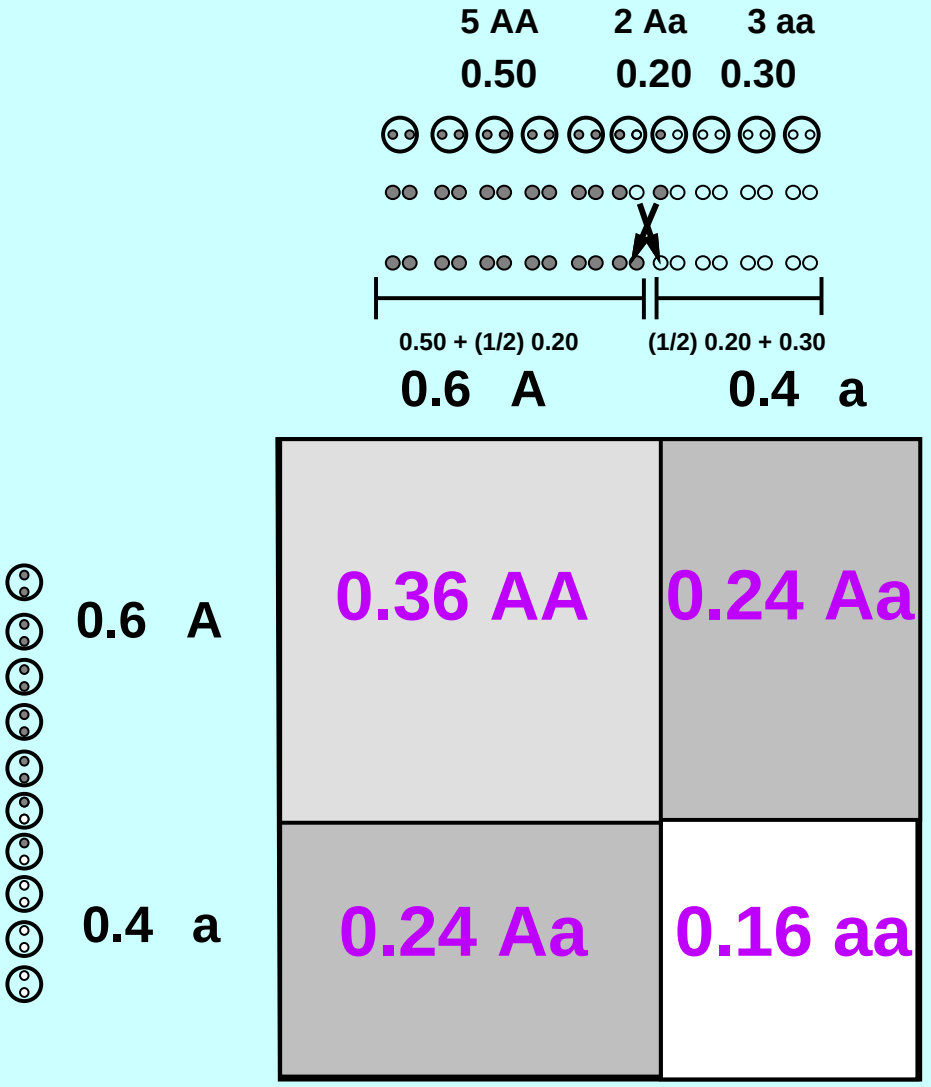
A Hardy-Weinberg calculation



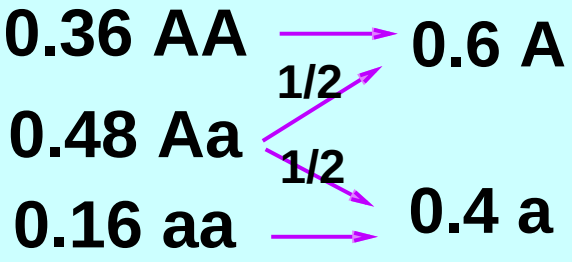
A Hardy-Weinberg calculation



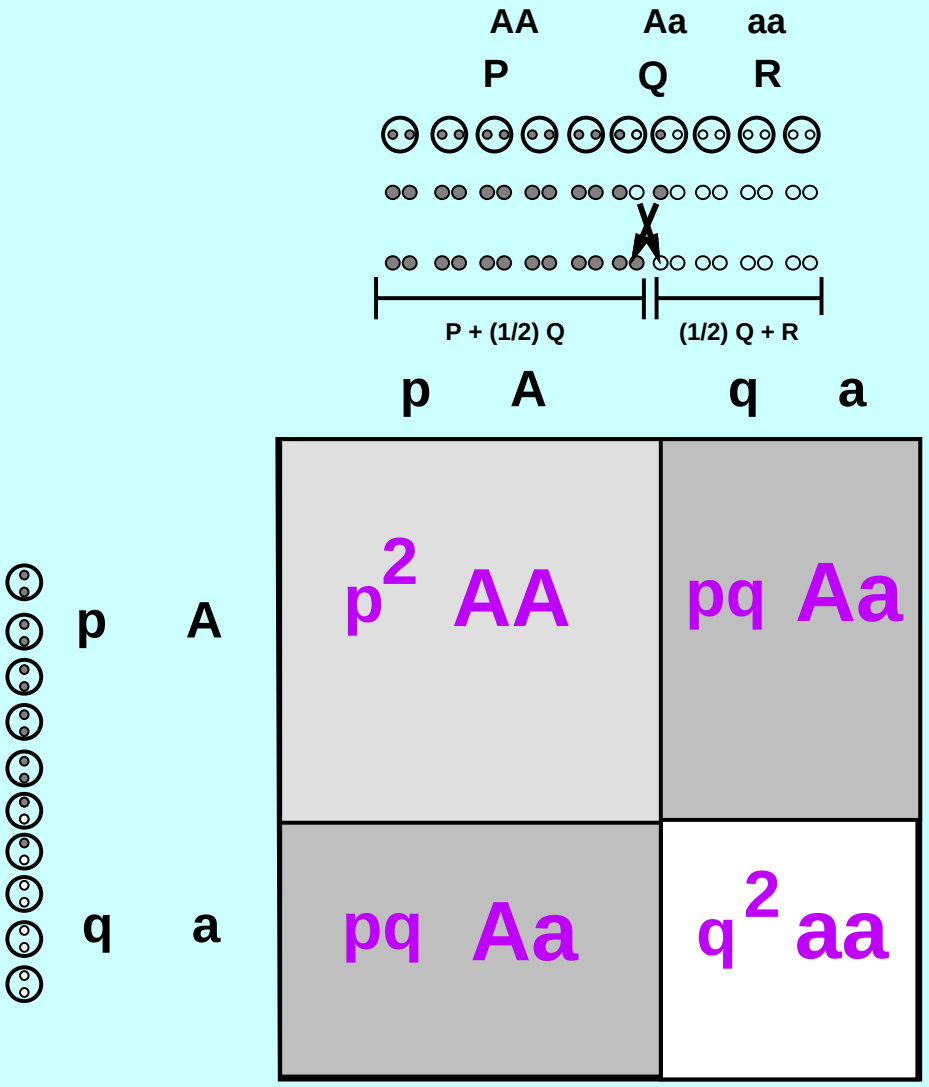
A Hardy-Weinberg calculation



Result:



Hardy-Weinberg mathematics



Result:

p^2 AA → p A

2pq Aa → p A (1/2) / q a (1/2)

q^2 aa → q a

Calculating the gene frequency (two ways)

Suppose that we have 200 individuals: 83 AA, 62 Aa, 55 aa

Method 1. Calculate what fraction of gametes bear A:

Genotype	Number	Genotype frequency		Fraction of gametes	
AA	83	0.415	all	0.57	A
Aa	62	0.31	1/2		
			1/2	0.43	a
aa	55	0.275	all		

Calculating the gene frequency (two ways)

Suppose that we have 200 individuals: 83 AA, 62 Aa, 55 aa

Method 2. Calculate what fraction of genes in the parents are A:

Genotype	Number	A's	a's
AA	83	166	0
Aa	62	62	62
aa	55	0	110
		228	172

228 + 172 = 400

$$\frac{228}{400} = 0.57 \quad \text{A}$$

$$\frac{172}{400} = 0.43 \quad \text{a}$$

A numerical example of natural selection

Genotypes:	AA	Aa	aa	
Relative fitnesses	1	1	0.7	(viabilities)
Initial gene frequency of A $= 0.2$				
(newborns)	0.04	0.32	0.64	
	$\times 1$	$\times 1$	$\times 0.7$	
Survivors:	0.04	$+ 0.32$	$+ 0.448$	$= 0.808$

Genotype frequencies among the survivors (divide by the total)

	0.0495	0.396	0.554

gene frequency

A :	$0.0495 + 0.5 \times 0.396$	$= 0.2475$
a :	$0.554 + 0.5 \times 0.396$	$= 0.7525$

genotype frequencies (among newborns)

0.0613	0.3725	0.5663

The algebra of natural selection

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Relative fitnesses	w_{AA}	w_{Aa}	w_{aa}
After selection	$p^2 w_{AA}$	$2pq\, w_{Aa}$	$q^2 w_{aa}$

$$\text{Sum} = p^2 w_{AA} + 2pq\, w_{Aa} + q^2 w_{aa}$$

Survivors: $\frac{p^2 w_{AA}}{\text{Sum}}$ $\frac{2pq\, w_{Aa}}{\text{Sum}}$ $\frac{q^2 w_{aa}}{\text{Sum}}$

$$p' = \frac{p^2 w_{AA} + pq\, w_{Aa}}{p^2 w_{AA} + 2pq\, w_{Aa} + q^2 w_{aa}}$$

$$p' = \frac{p(p\, w_{AA} + q\, w_{Aa})}{p^2 w_{AA} + 2pq\, w_{Aa} + q^2 w_{aa}}$$

$$p' = \frac{p\, \bar{w}_A}{\bar{w}}$$

Is weak selection effective?

Suppose (relative) fitnesses are:

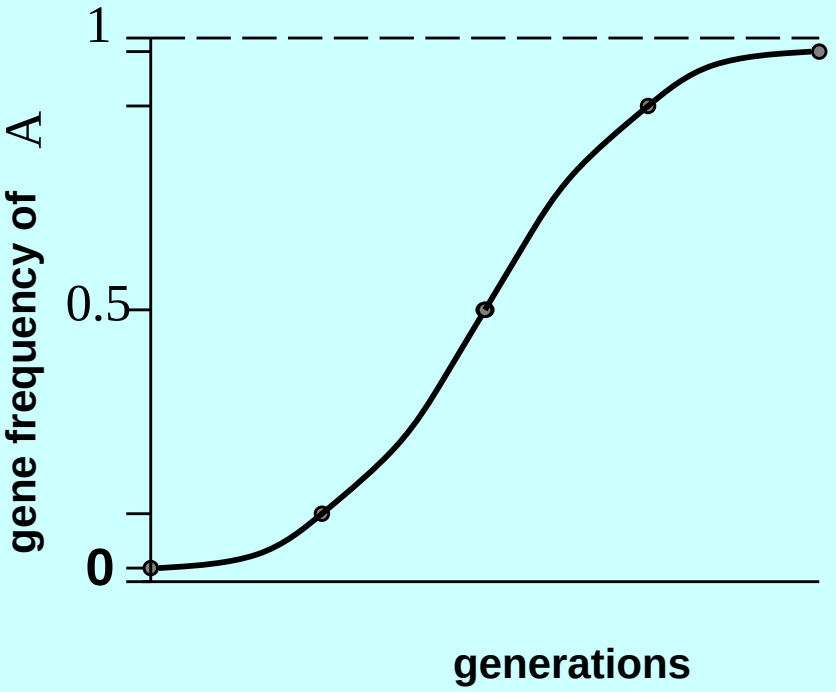
AA	Aa	aa
$(1+s)^2$	$1+s$	1

x (1+s) x (1+s)

So in this example each change of a to A multiplies the fitness by (1+s), so that it increases it by a fraction s.

The time for gene frequency change, in generations, turns out to be:

	change of gene frequencies			
s	0.01 − 0.1	0.1 − 0.5	0.5 − 0.9	0.9 − 0.99
1	3.46	3.17	3.17	3.46
0.1	25.16	23.05	23.05	25.16
0.01	240.99	220.82	220.82	240.99
0.001	2399.09	2198.02	2198.02	2399.09



An experimental selection curve

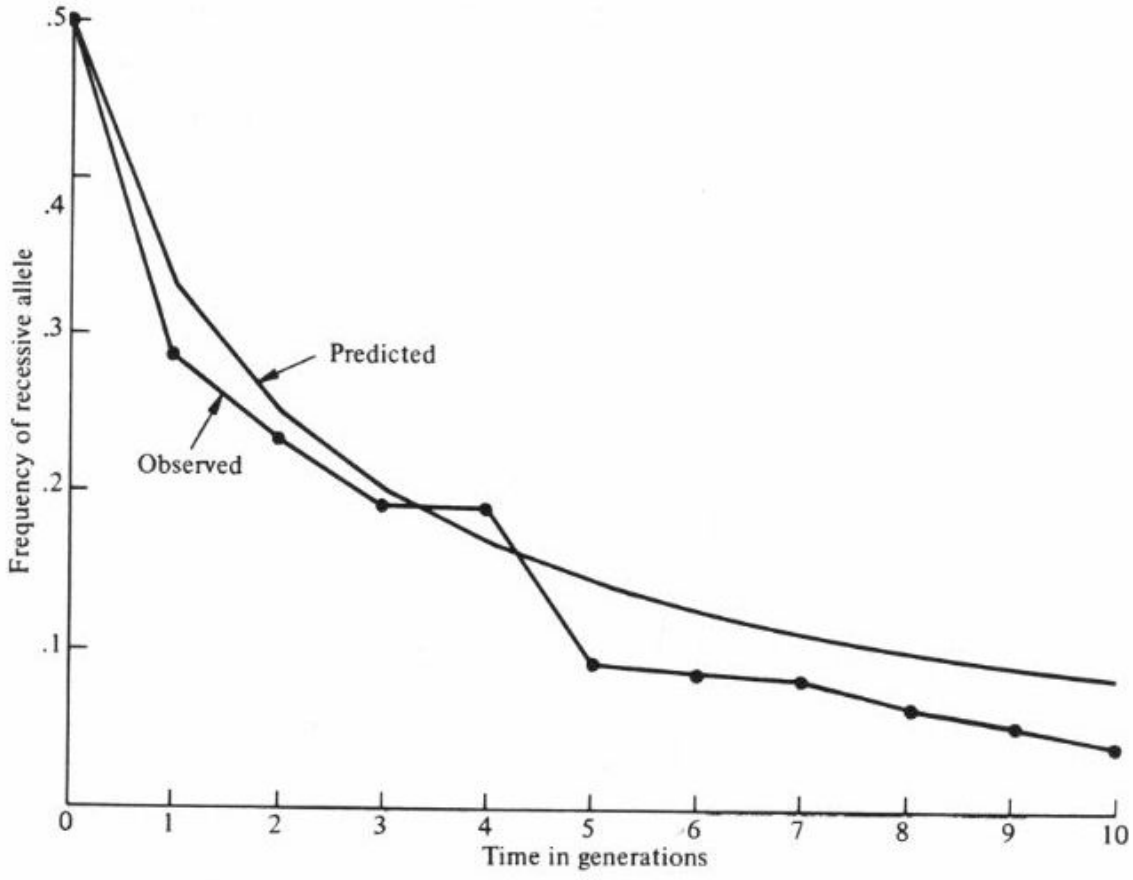
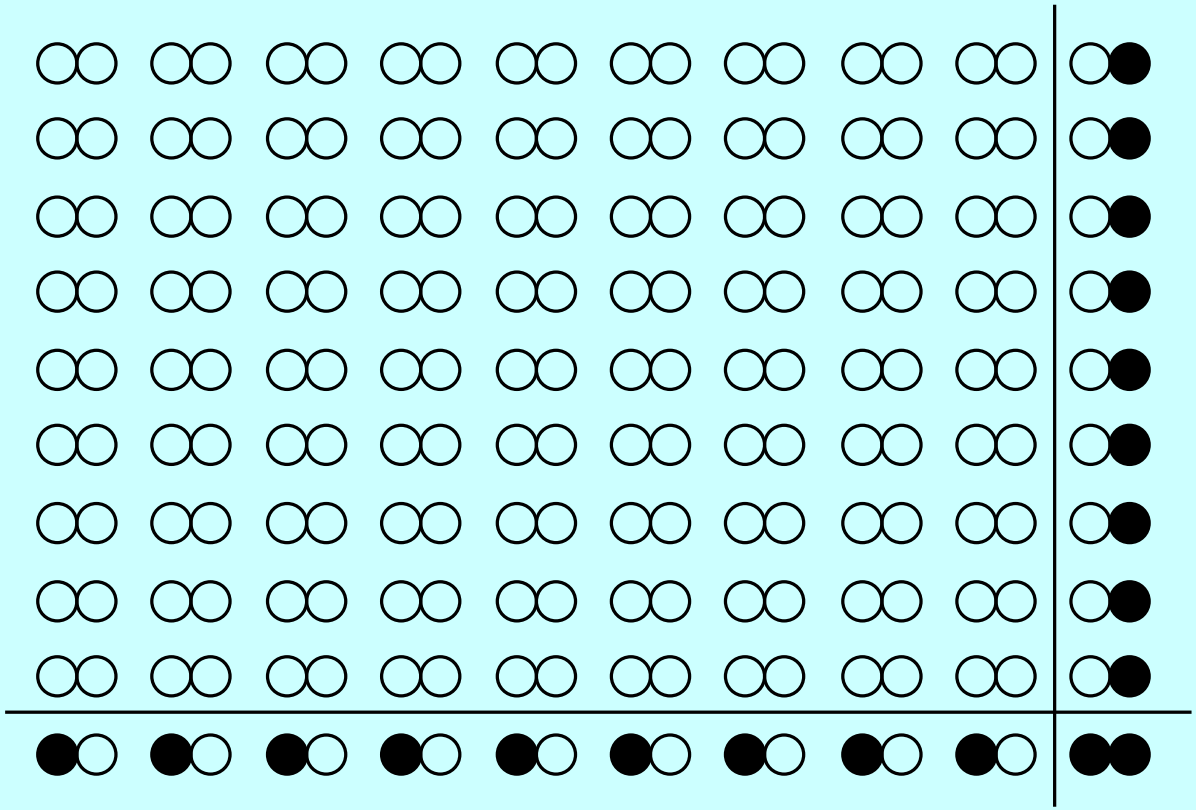


FIGURE 3.4. *Experiment illustrating selection against a recessive lethal gene. The frequency of the recessive allele is on the vertical axis, time in generations is on the horizontal axis.* [Data from B. Wallace (1963), The elimination of an autosomal lethal from an experimental population of *Drosophila melanogaster*, *Amer. Natur.* **97**: 65–66.]

Rare alleles occur mostly in heterozygotes



This shows a population in Hardy−Weinberg equilibrium

at gene frequencies of 0.9 A : 0.1 a

Genotype frequencies:

0.81 AA : 0.18 Aa : 0.01 aa

Note that of the 20 copies of a,

18 of them, or 18 / 20 = 0.9 of them are in Aa genotypes

Overdominance and polymorphism

AA	Aa	aa
$1-s$	1	$1-t$

when A is rare, most A's are in Aa, and most a's are in aa

The average fitness of A−bearing genotypes is then nearly 1

The average fitness of a−bearing genotypes is then nearly $1-t$

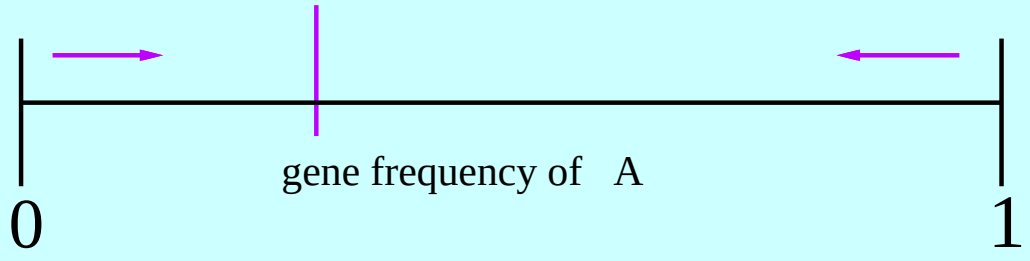
So A will increase in frequency when rare

when a is rare, most a's are in Aa, and most A's are in AA

The average fitness of a−bearing genotypes is then nearly 1

The average fitness of A−bearing genotypes is then nearly $1-s$

So a will increase in frequency when rare



Underdominance and unstable equilibrium

AA	Aa	aa
1+s	1	1+t

when A is rare, most A's are in Aa, and most a's are in aa

The average fitness of A−bearing genotypes is then nearly 1

The average fitness of a−bearing genotypes is then nearly 1+t

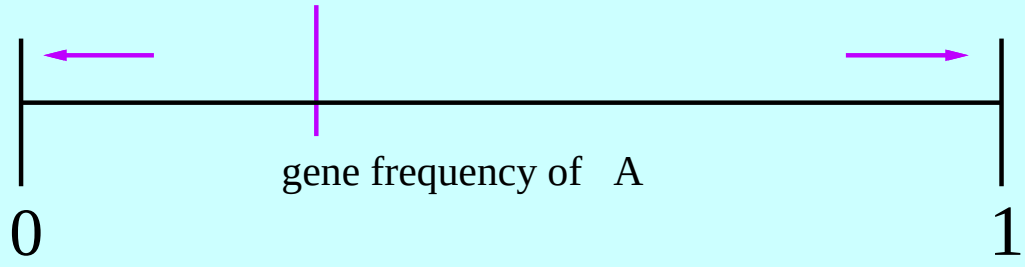
So A will decrease in frequency when rare

when a is rare, most a's are in Aa, and most A's are in AA

The average fitness of a−bearing genotypes is then nearly 1

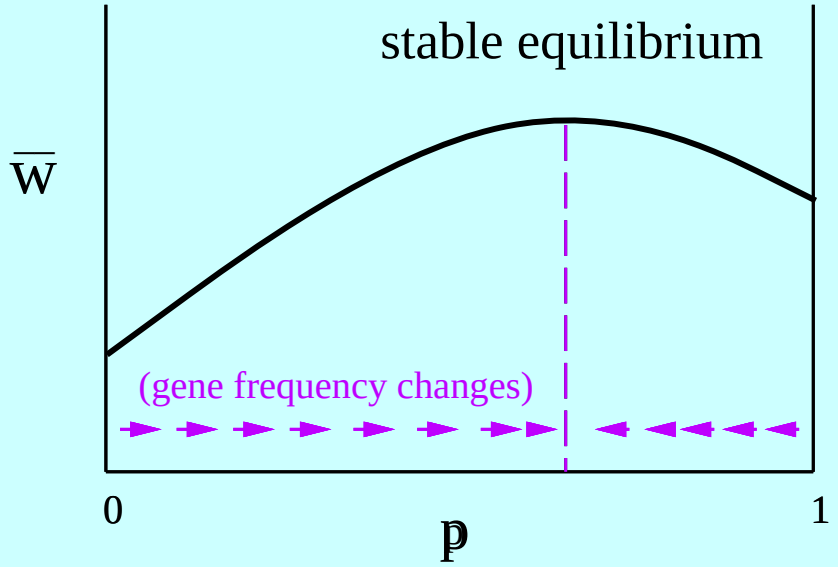
The average fitness of A−bearing genotypes is then nearly 1+s

So a will decrease in frequency when rare

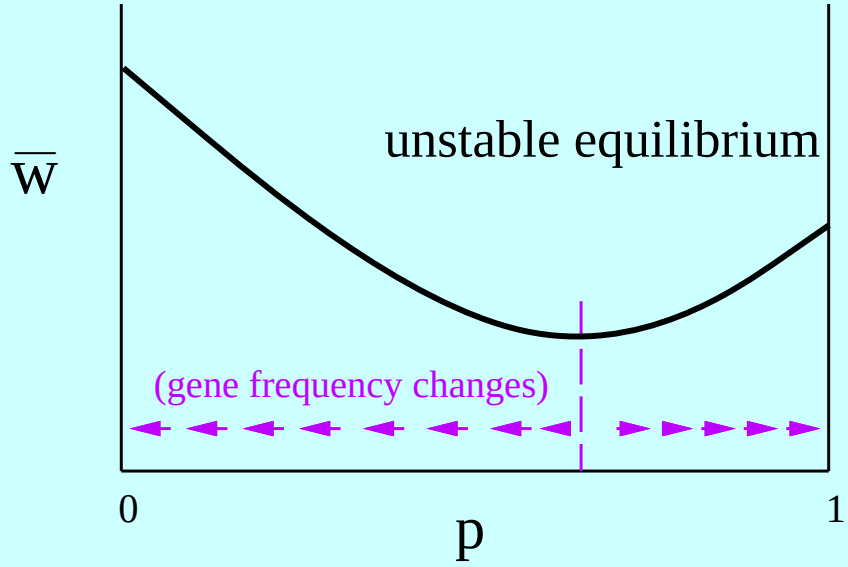


Fitness surfaces (adaptive landscapes)

Overdominance

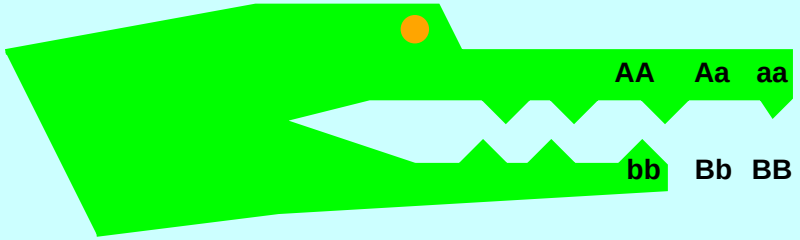


Underdominance



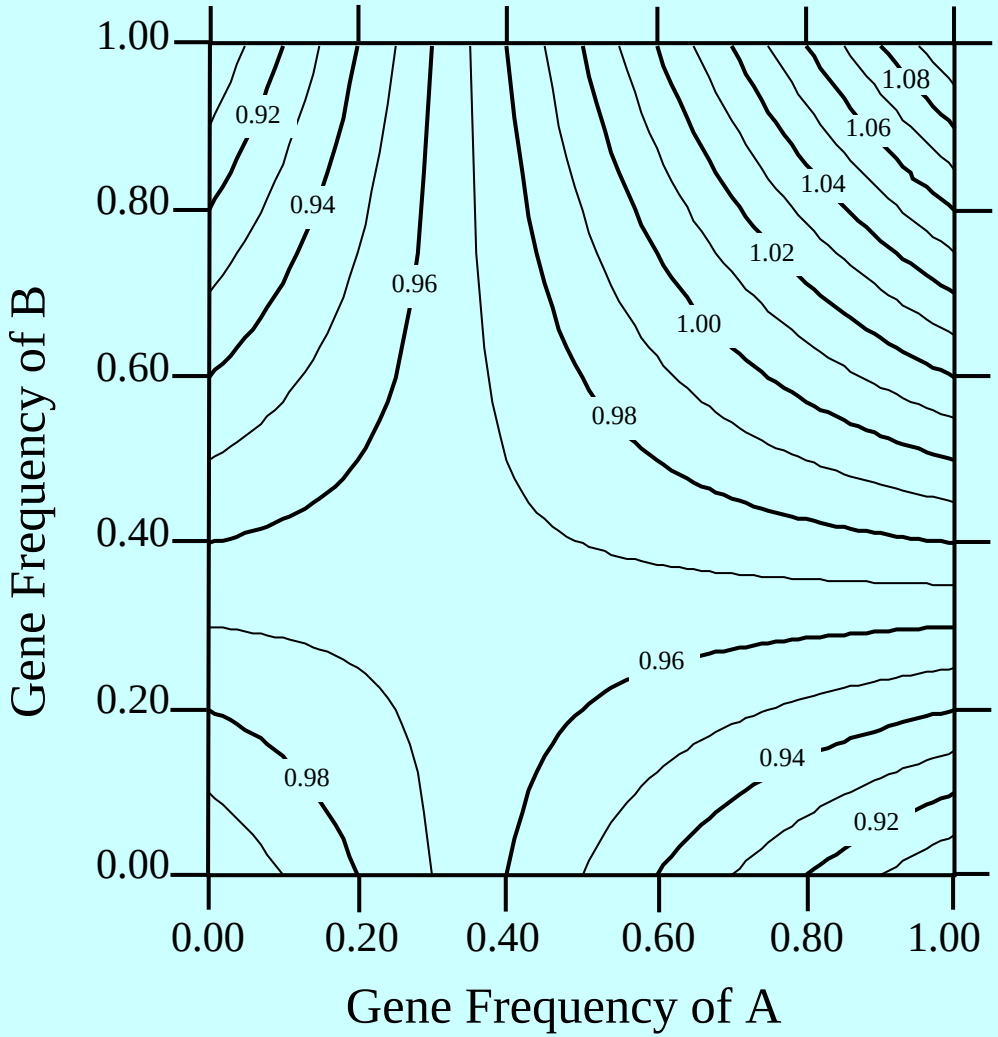
Is all for the best in this best of all possible worlds?

A case to consider: two interacting adaptations



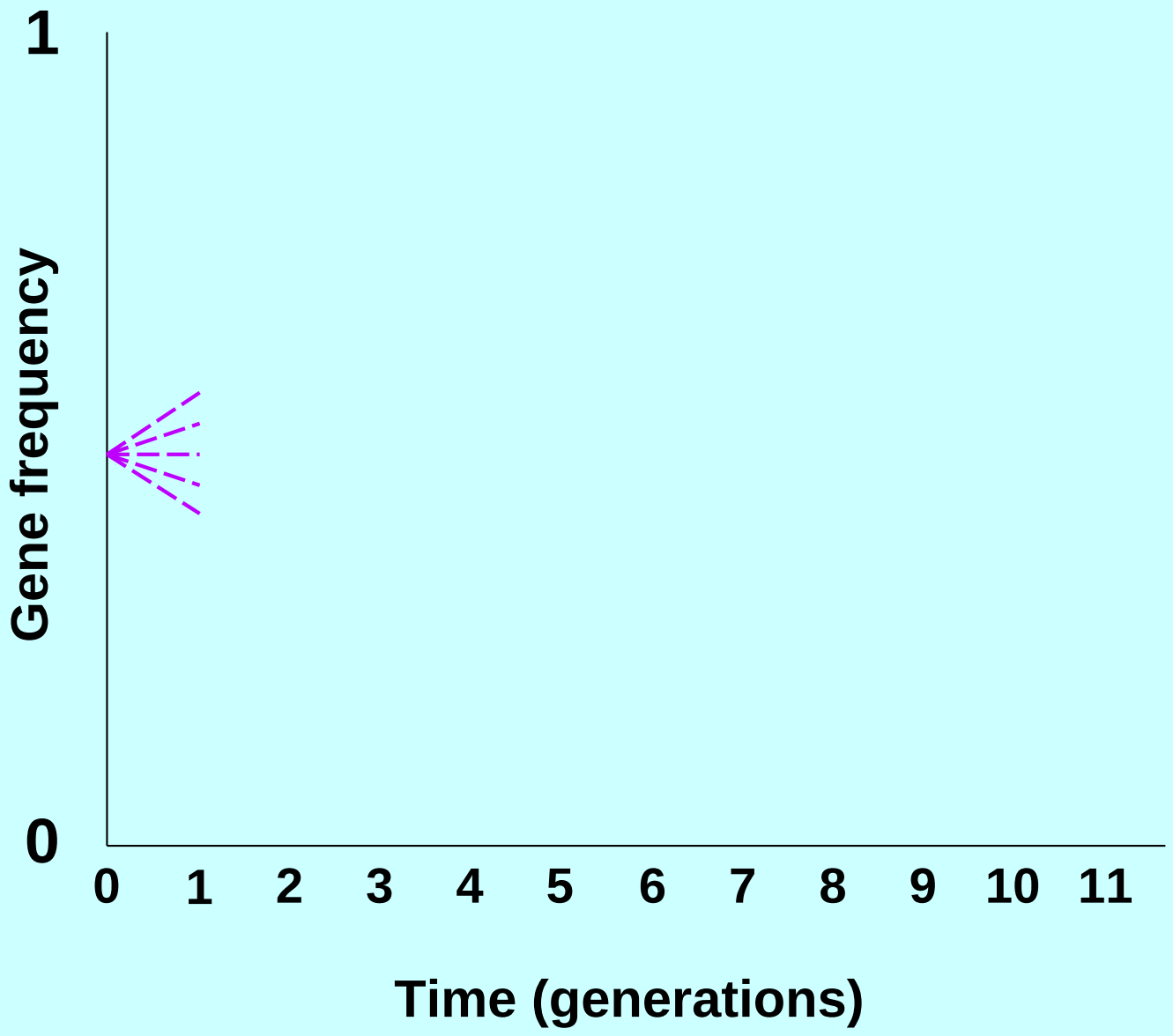
	BB	*Bb*	*bb*
AA	0.8	0.9	1.0
Aa	0.9	1.0	0.9
aa	1.0	0.9	0.8

A fitness surface (in a haploid case)

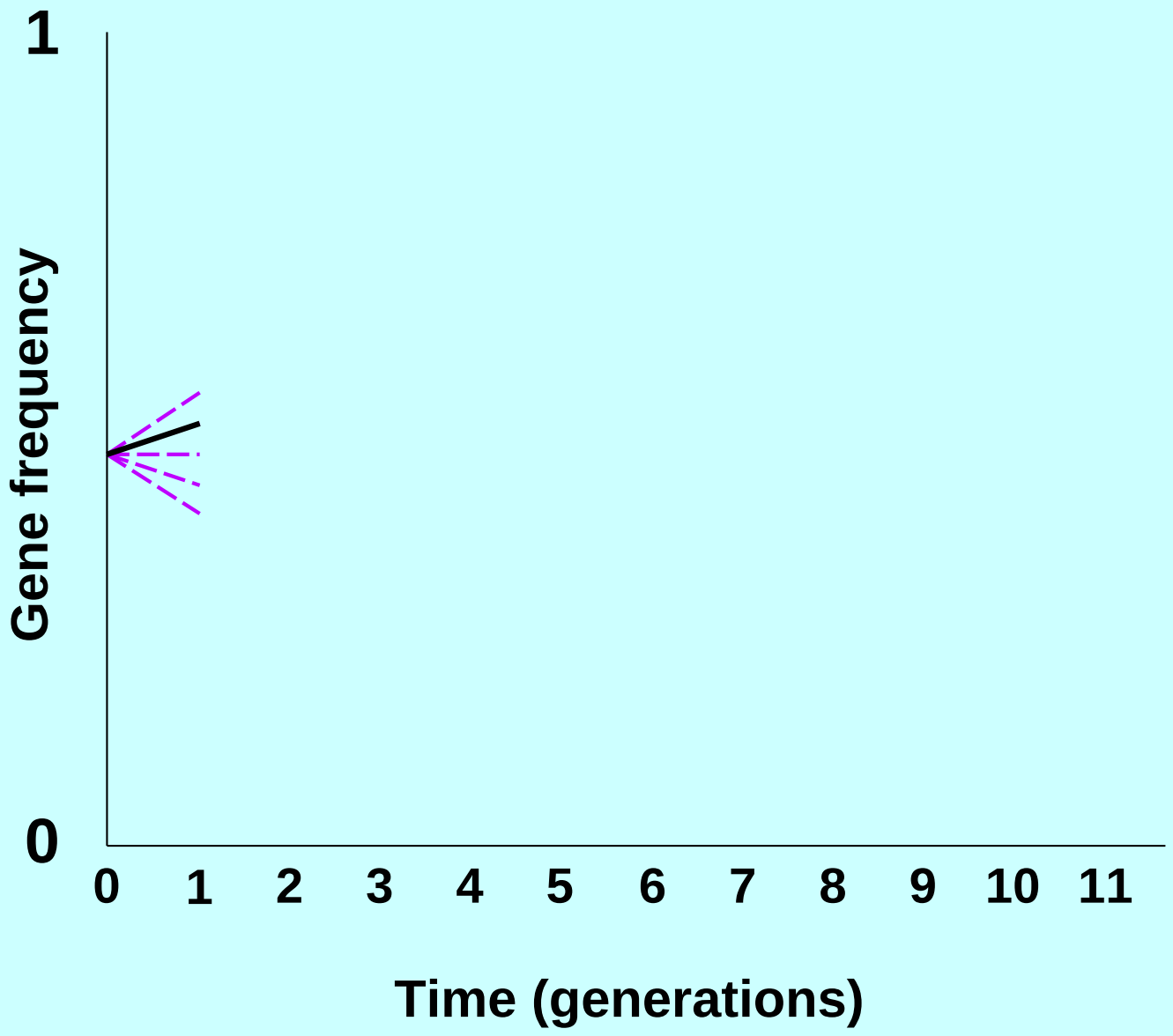


AB	1.1
Ab	0.9
aB	0.9
ab	1

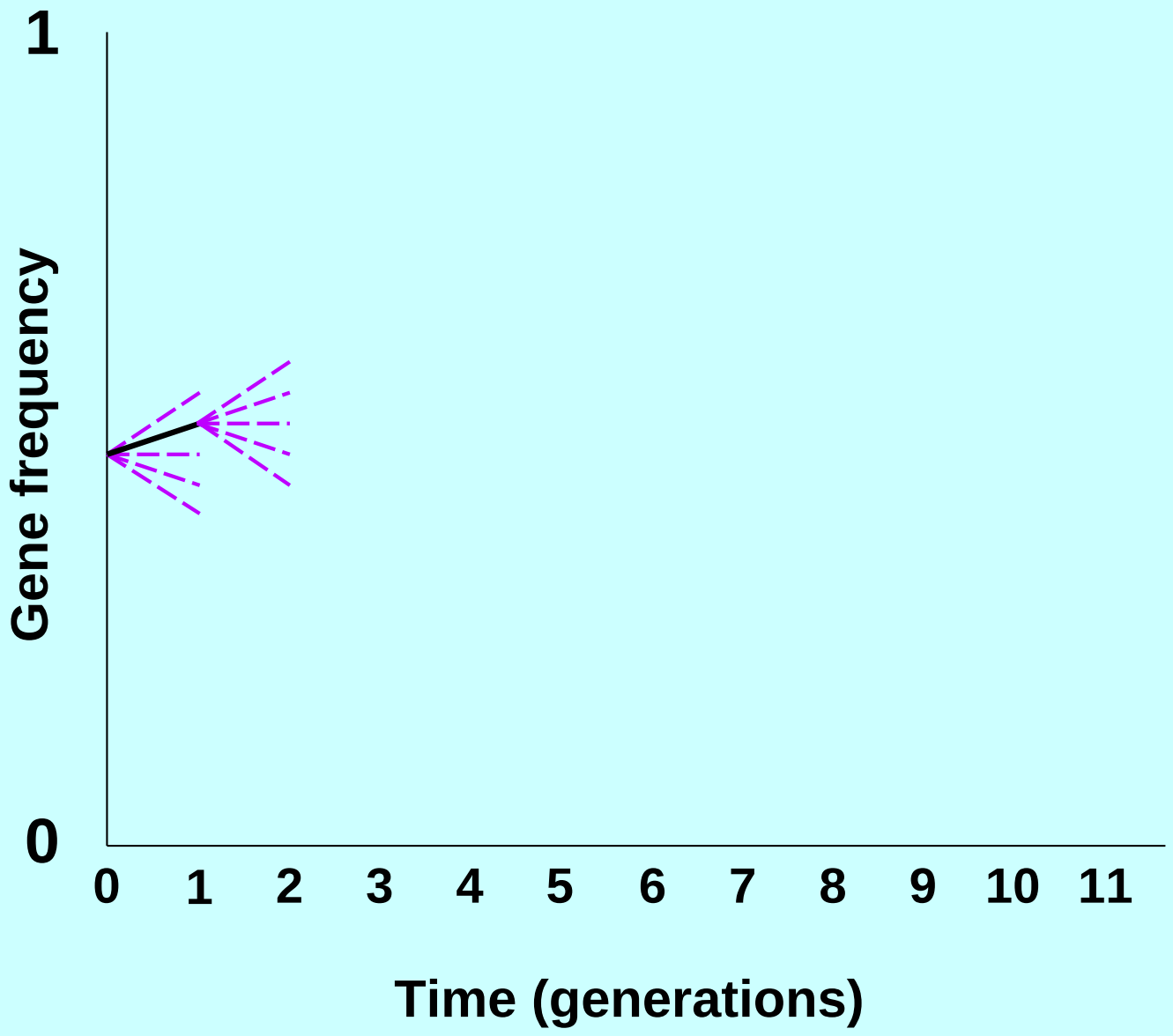
Genetic drift



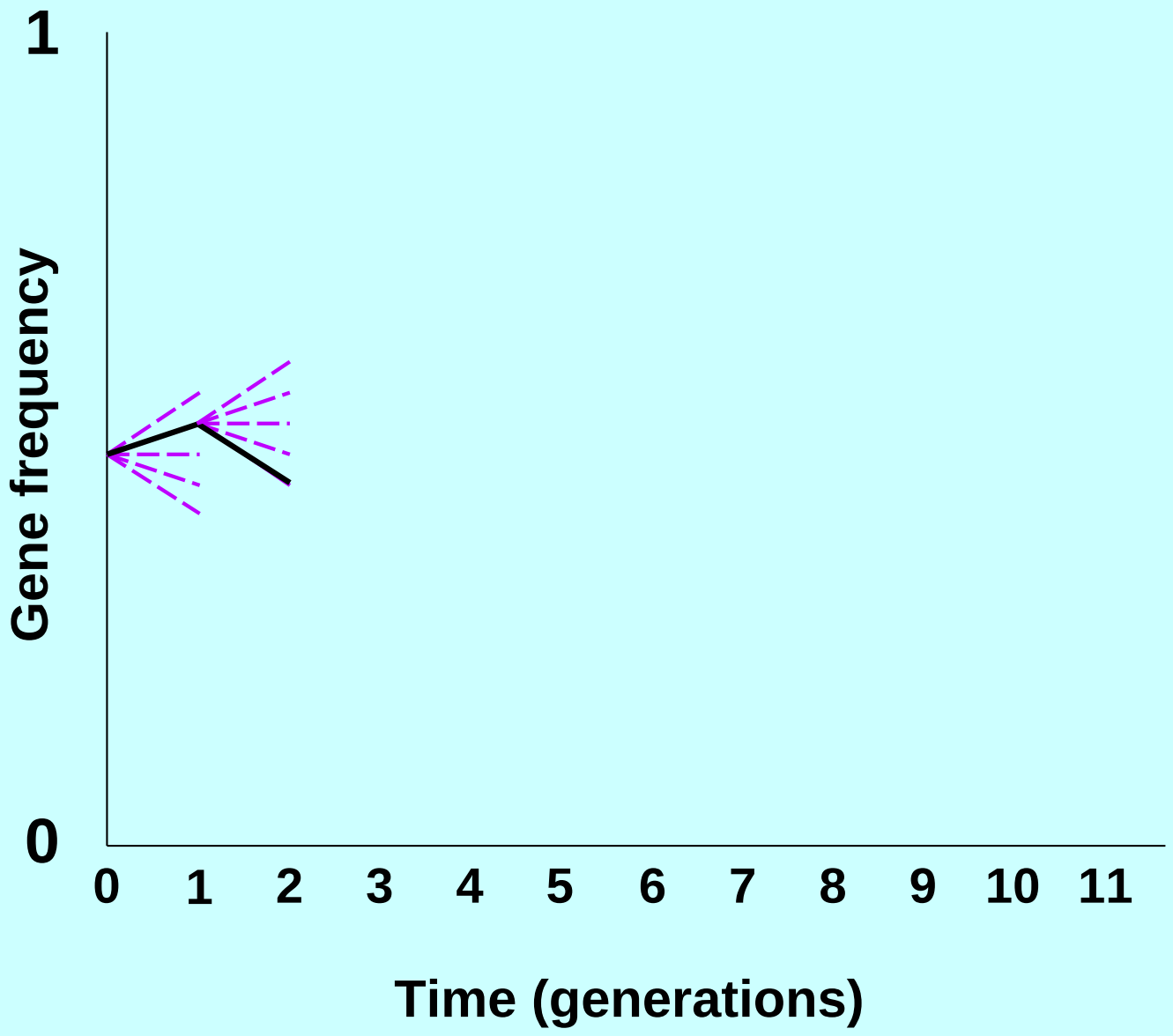
Genetic drift



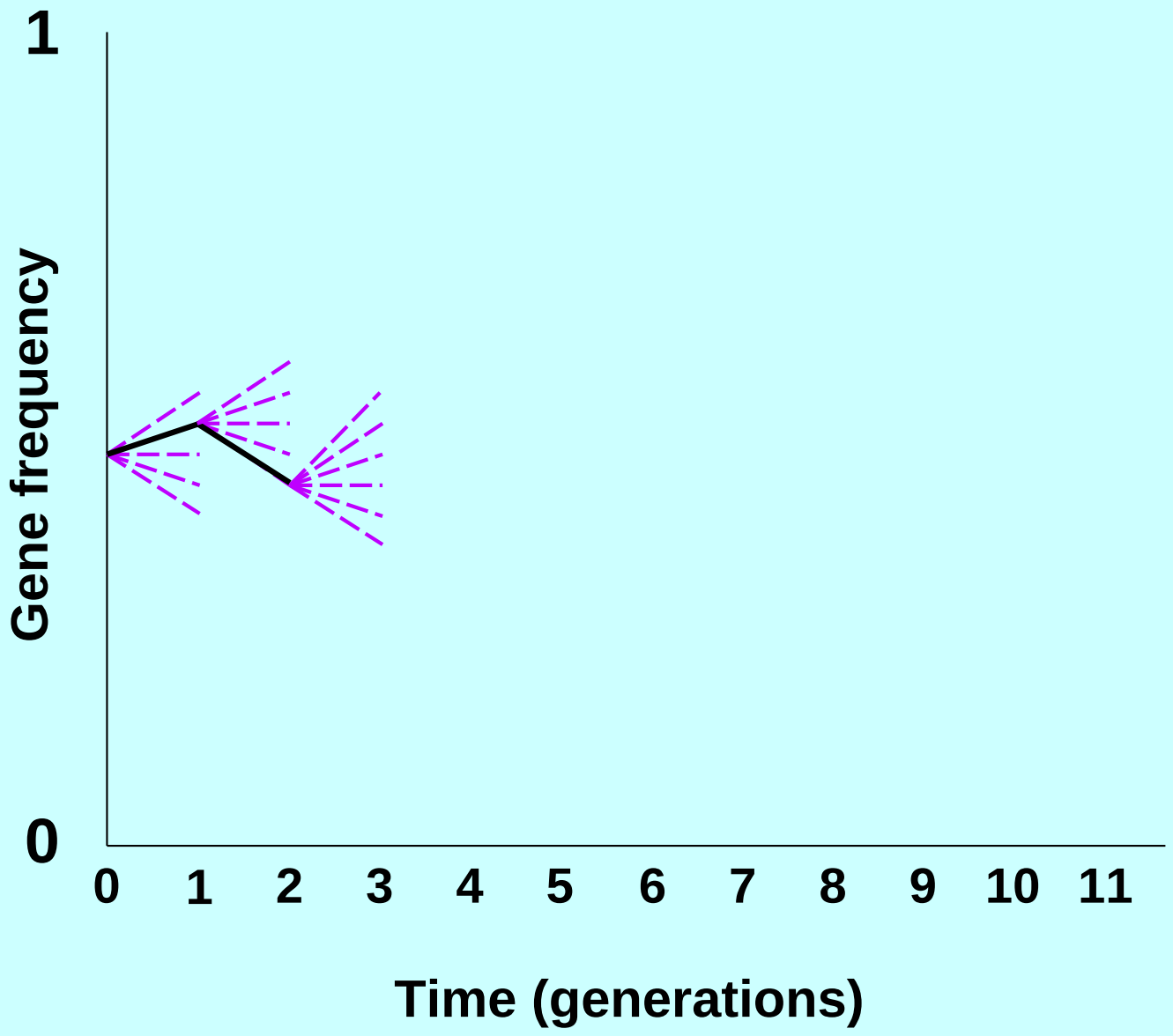
Genetic drift



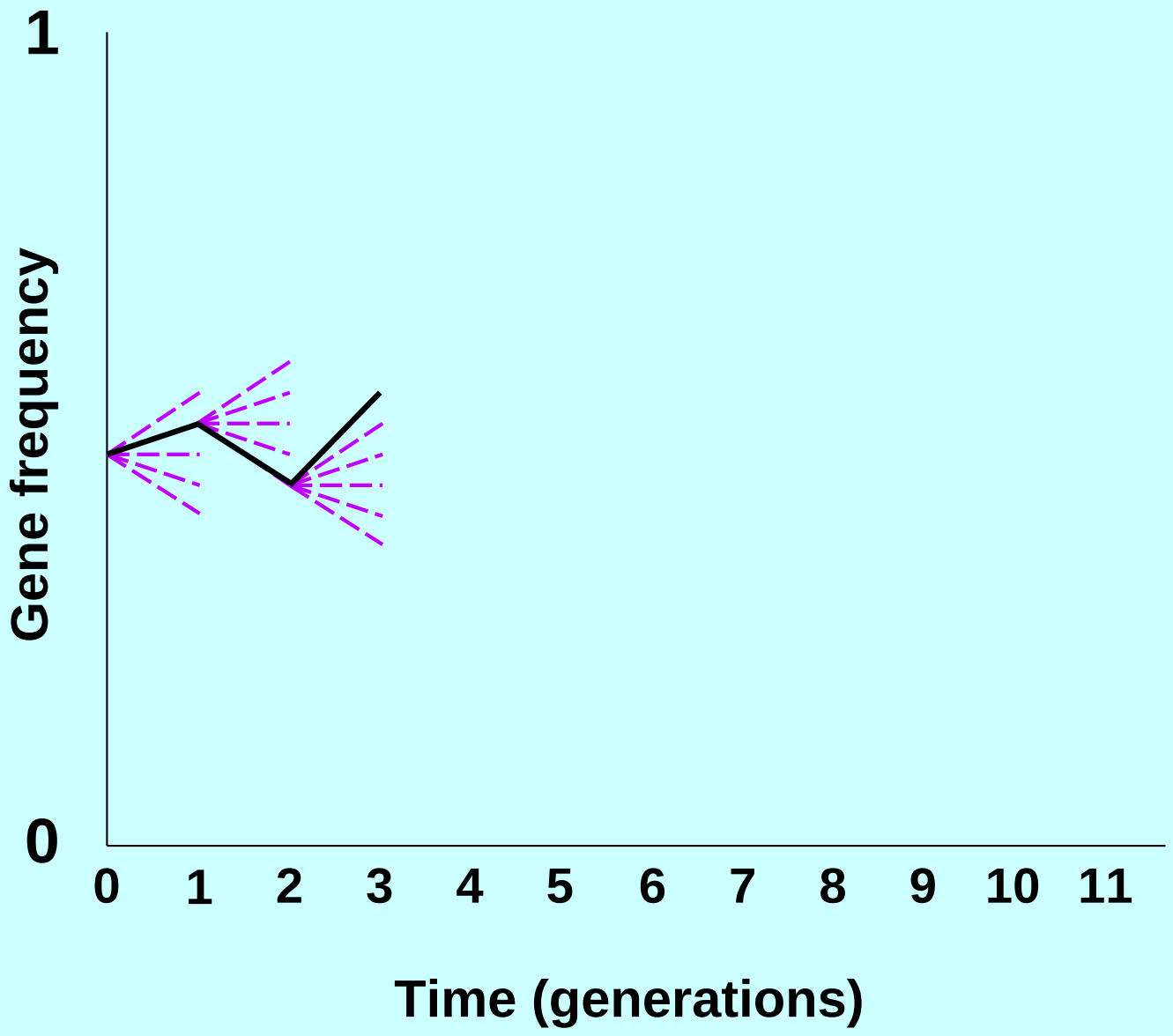
Genetic drift



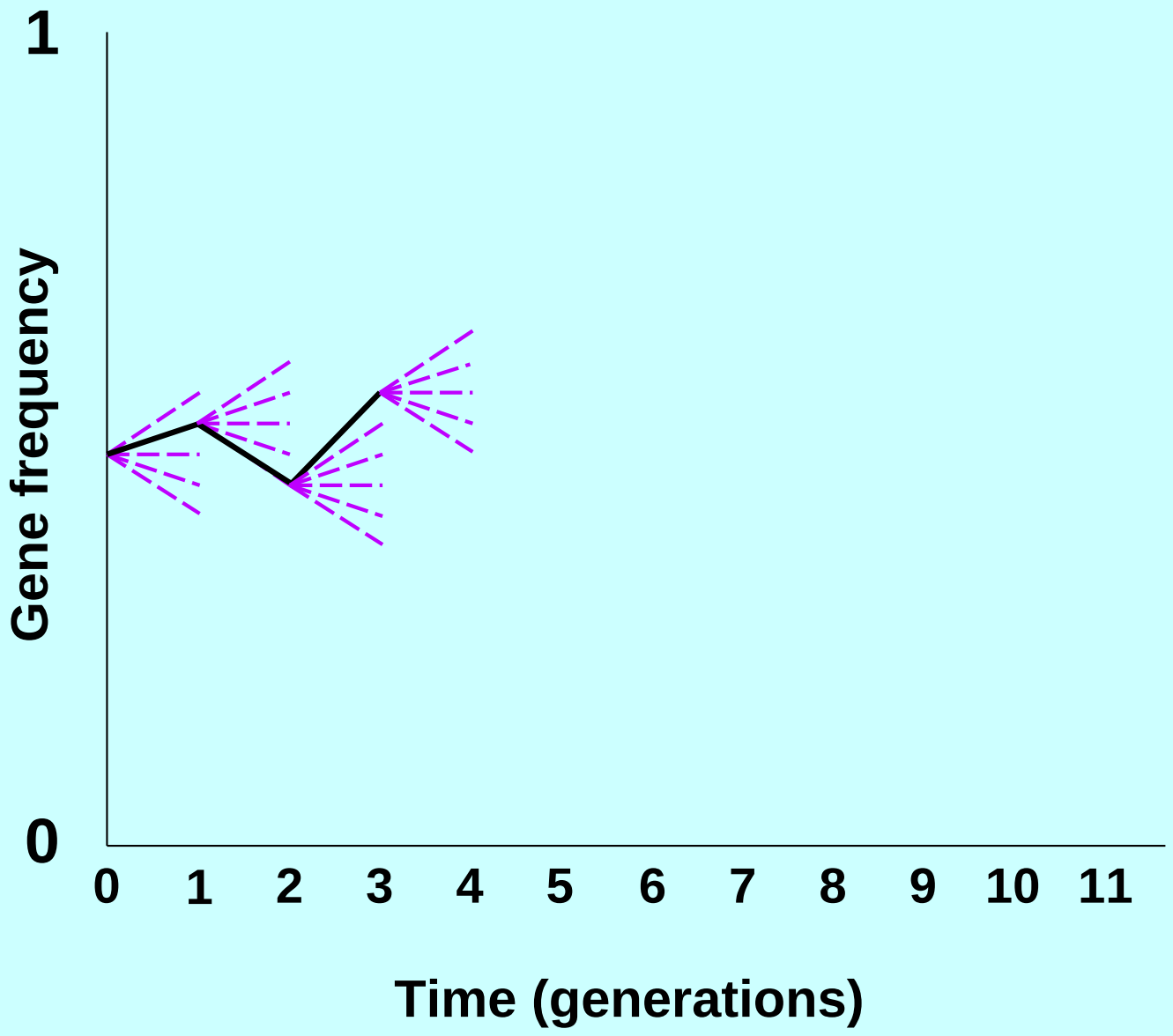
Genetic drift



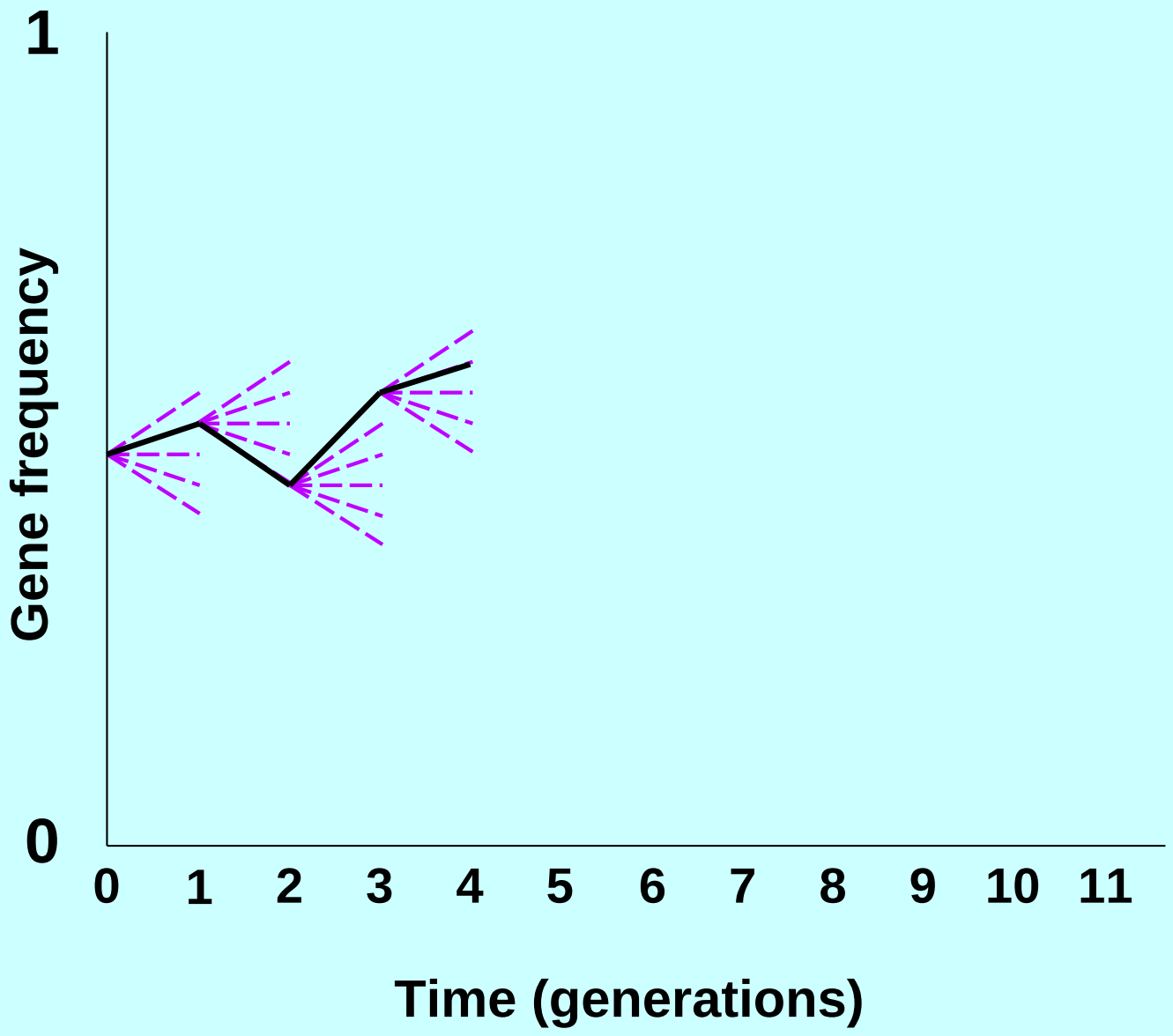
Genetic drift



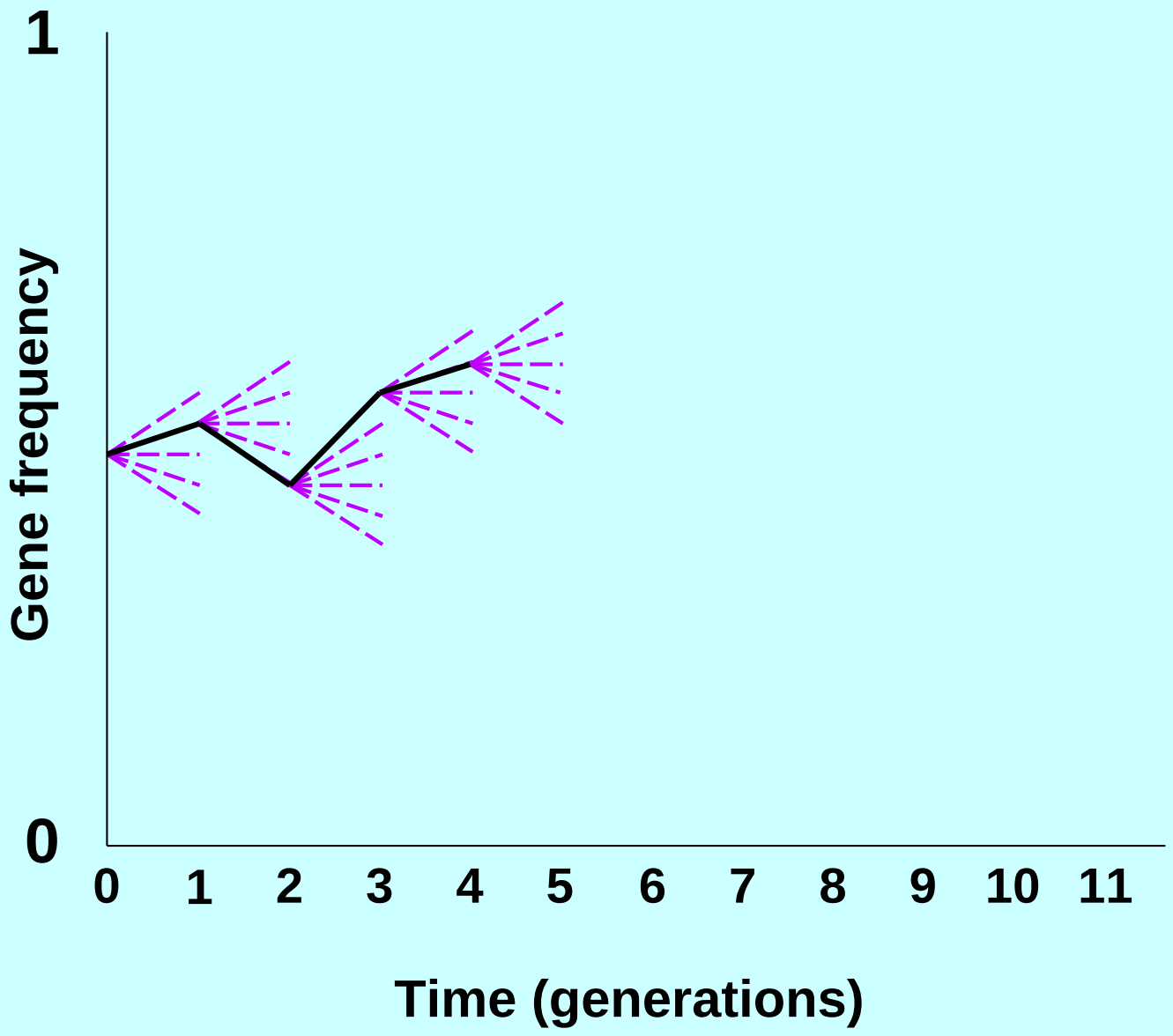
Genetic drift



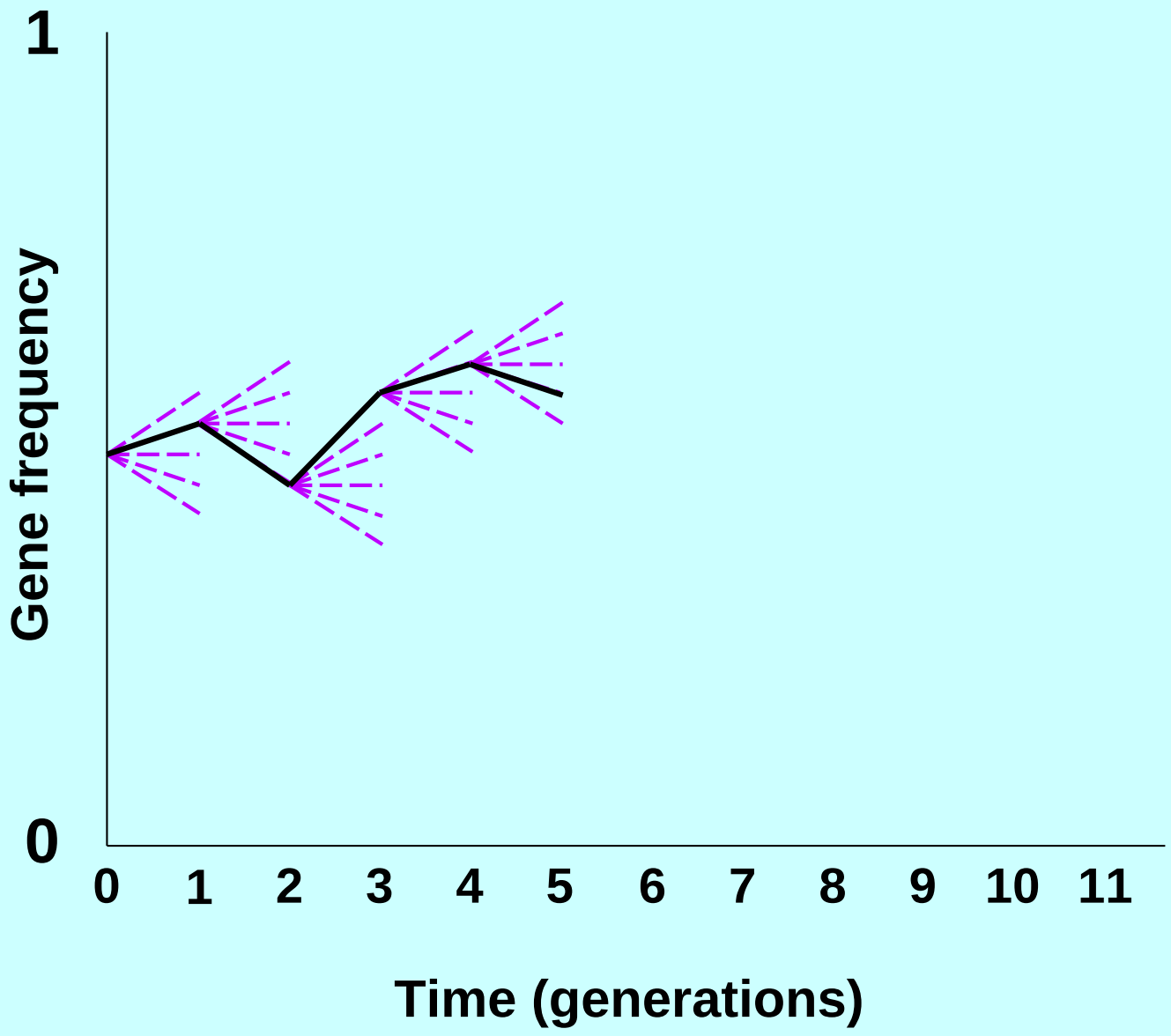
Genetic drift



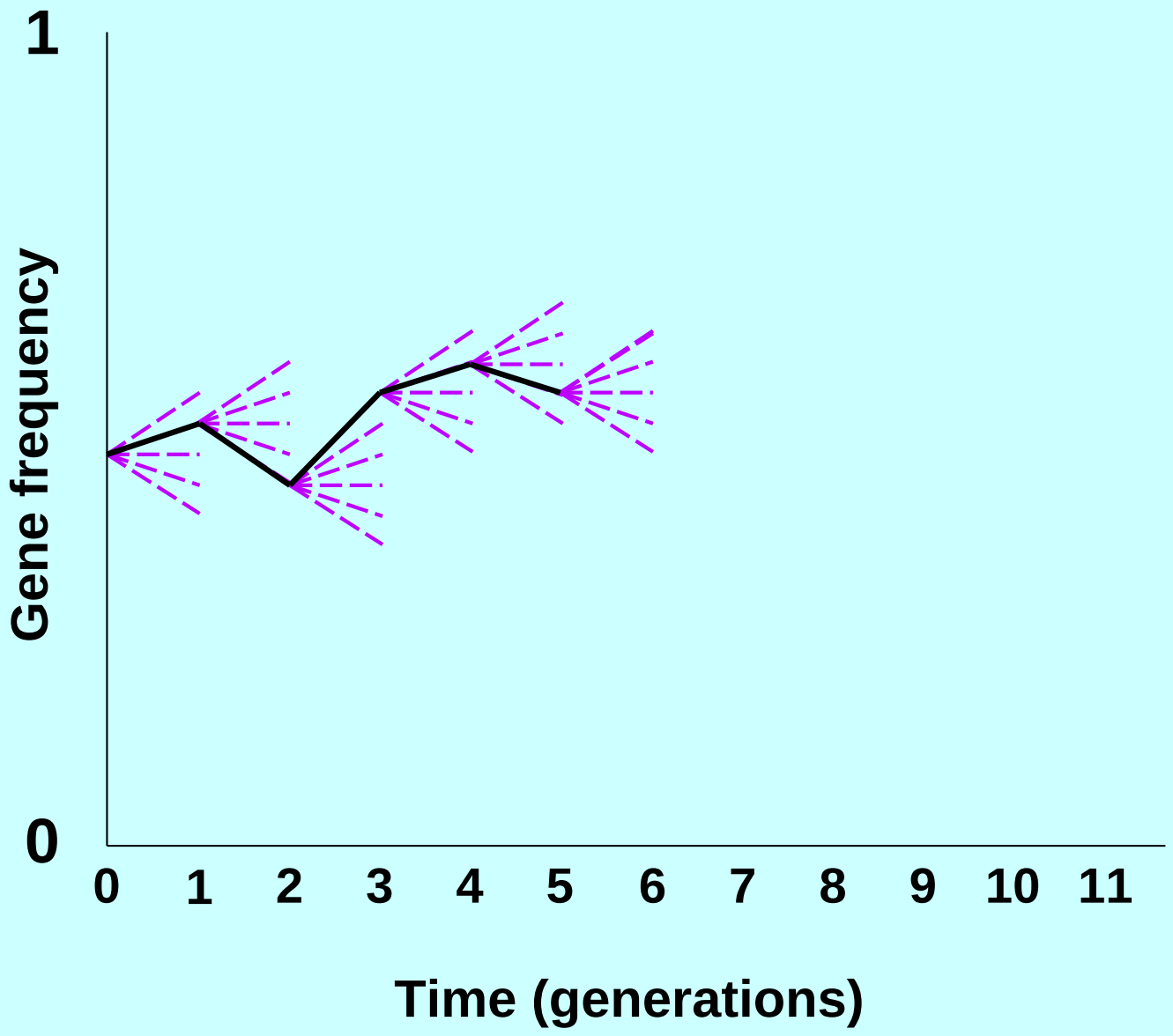
Genetic drift



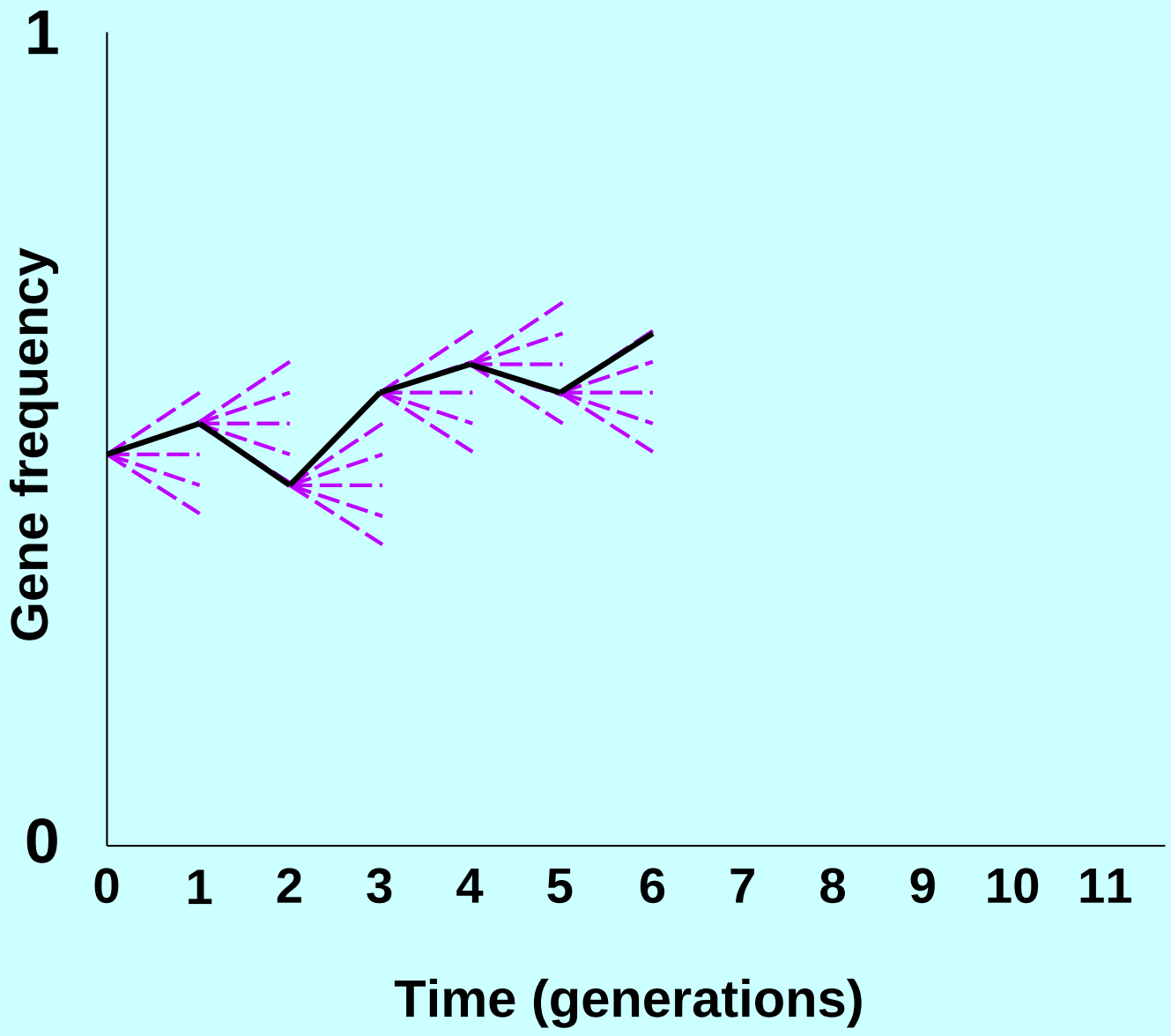
Genetic drift



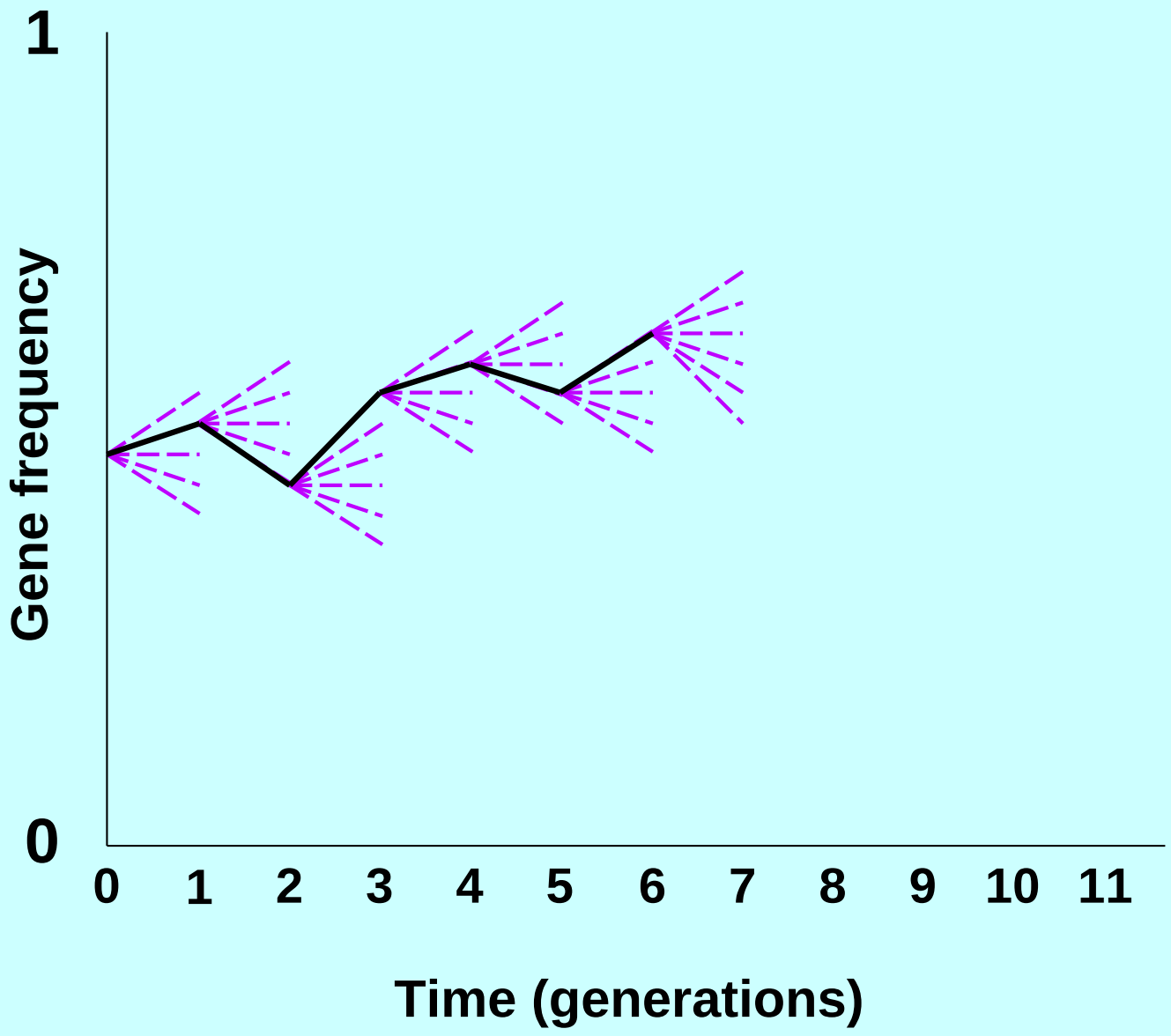
Genetic drift



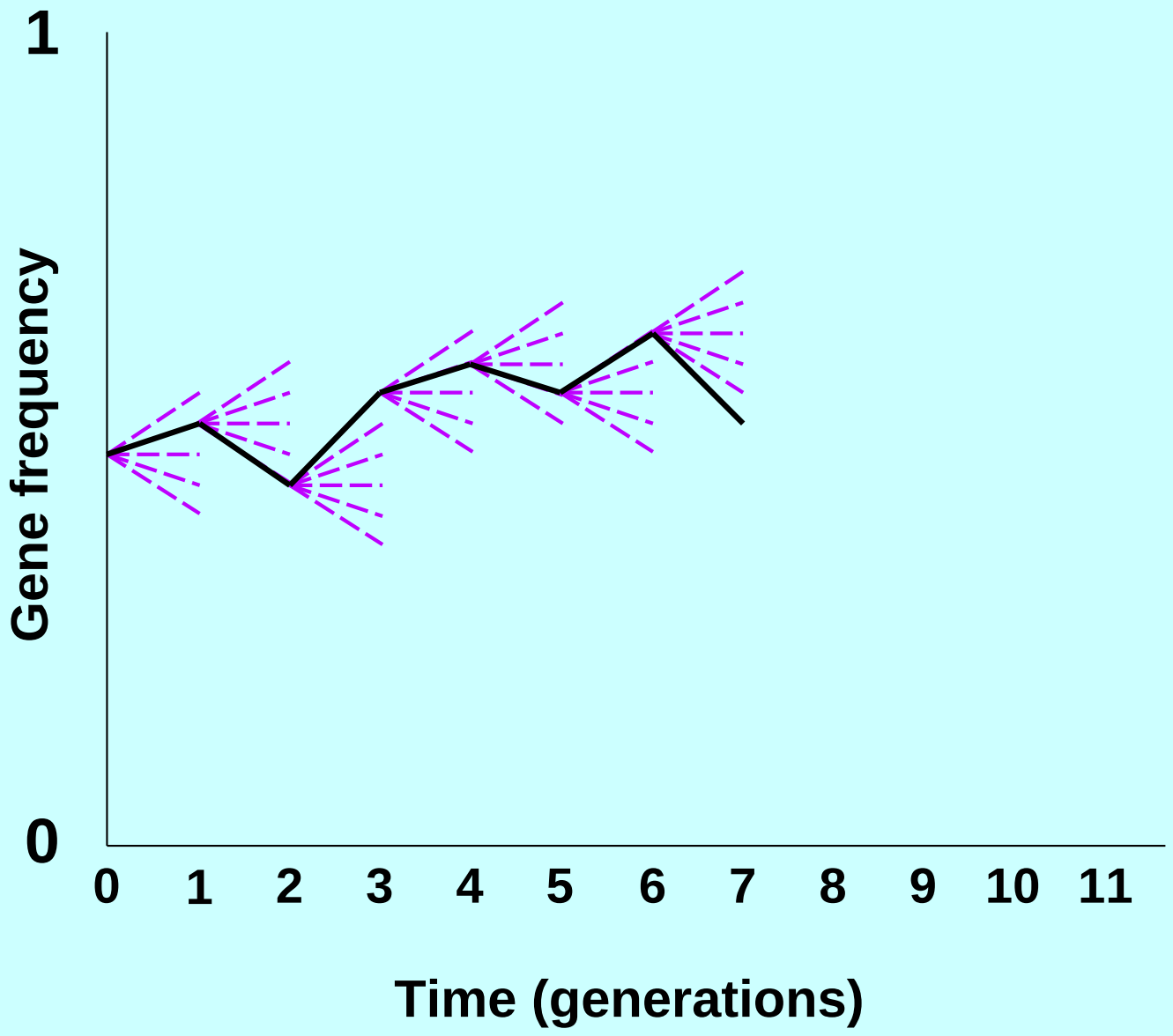
Genetic drift



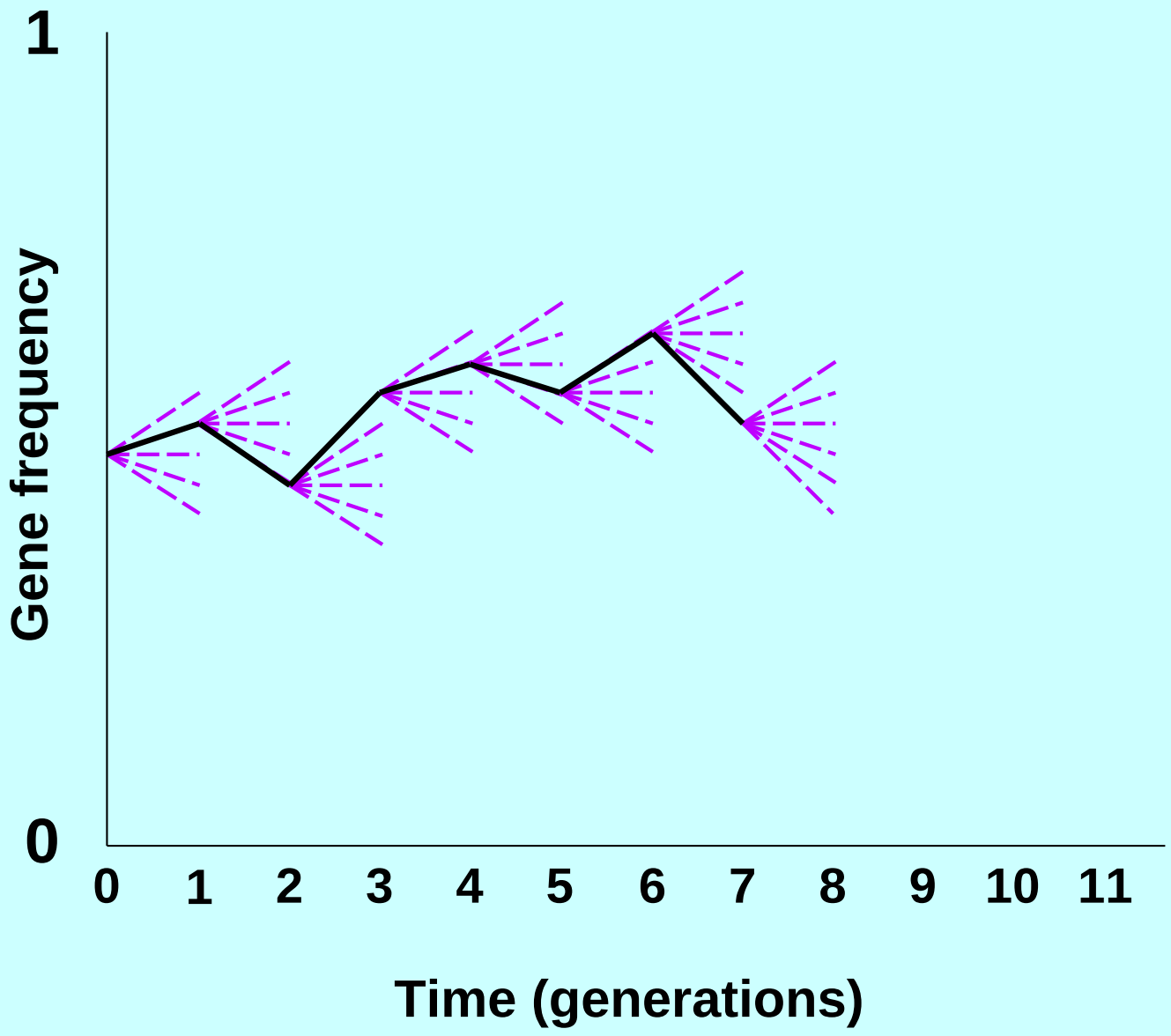
Genetic drift



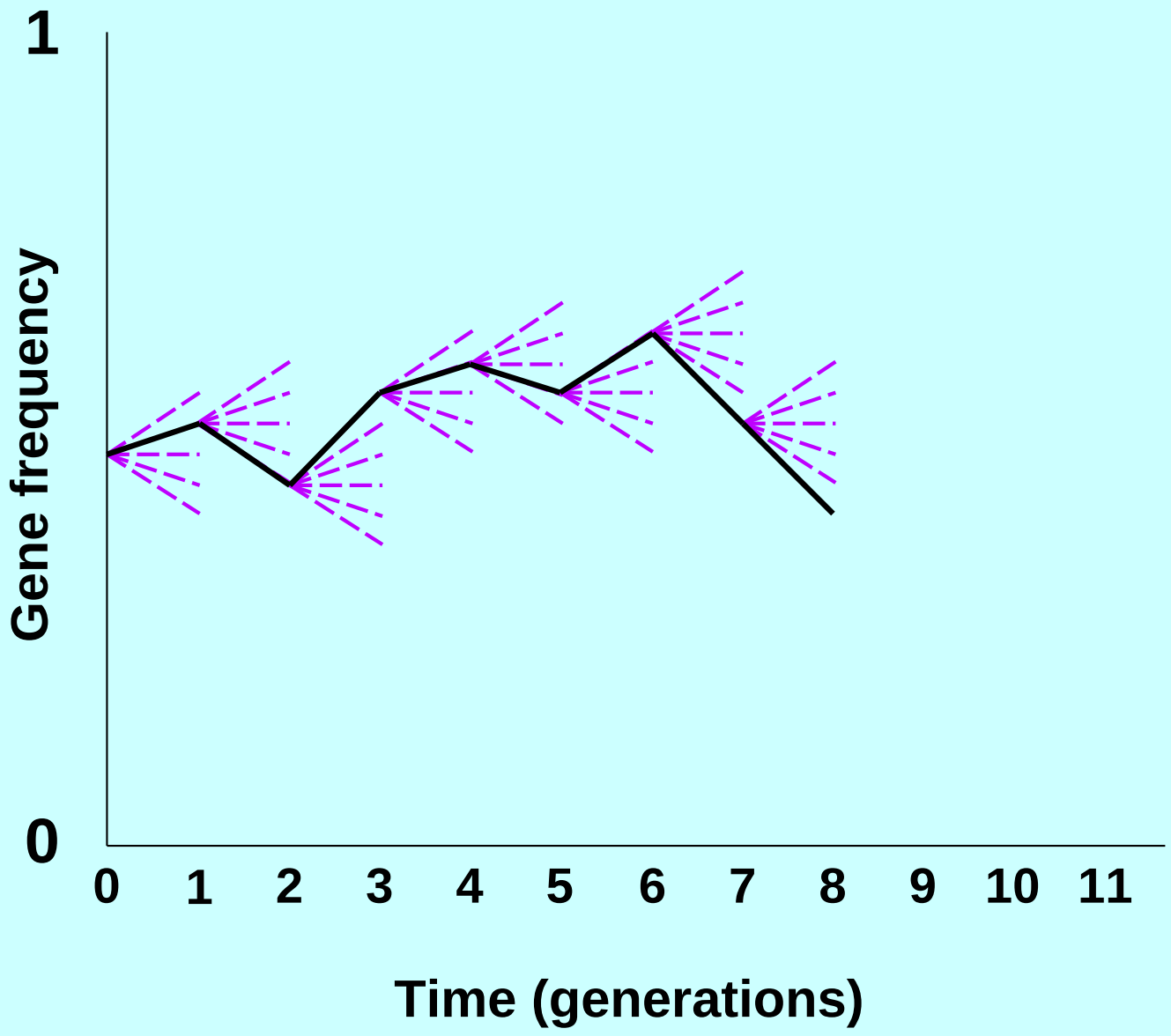
Genetic drift



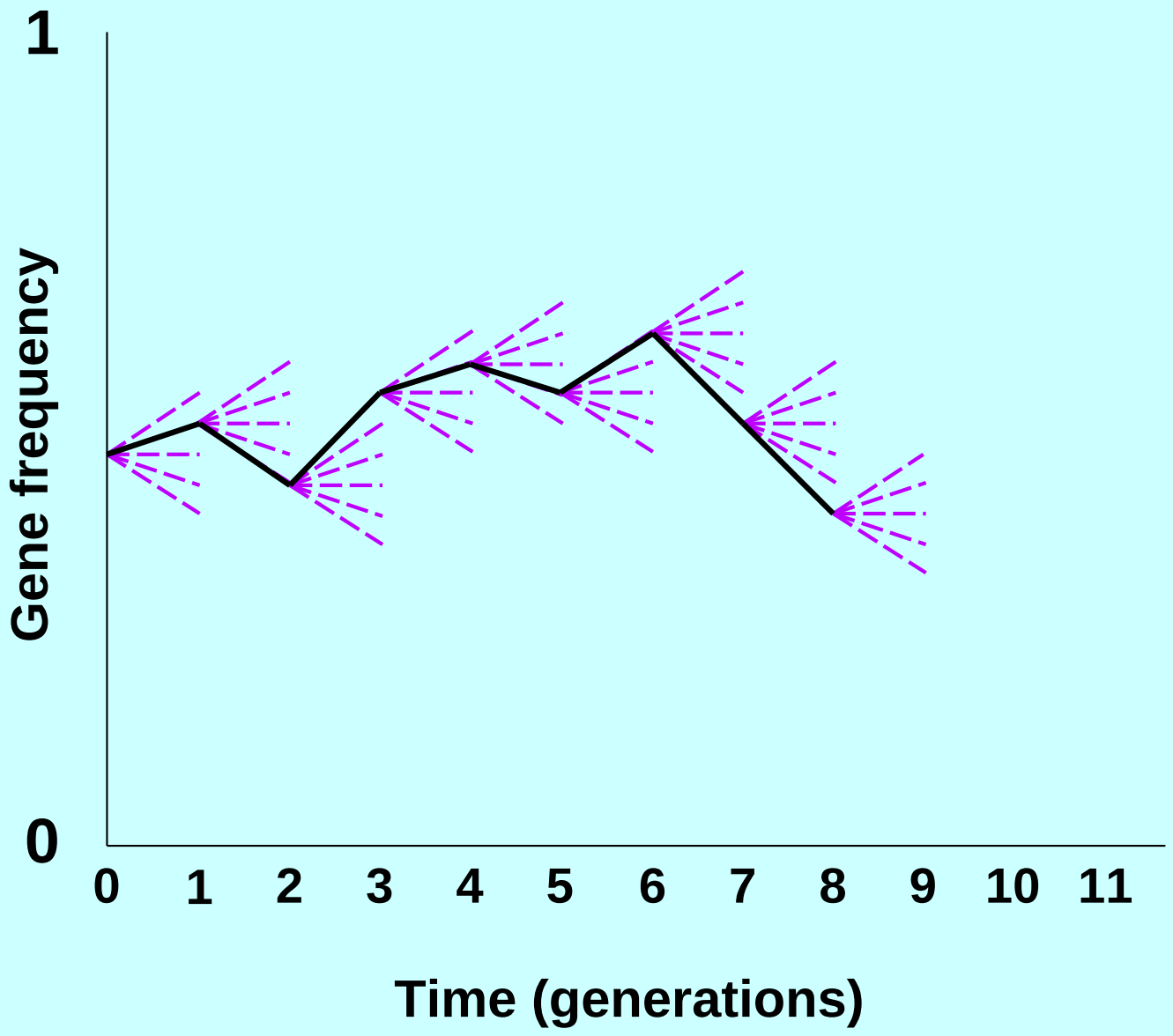
Genetic drift



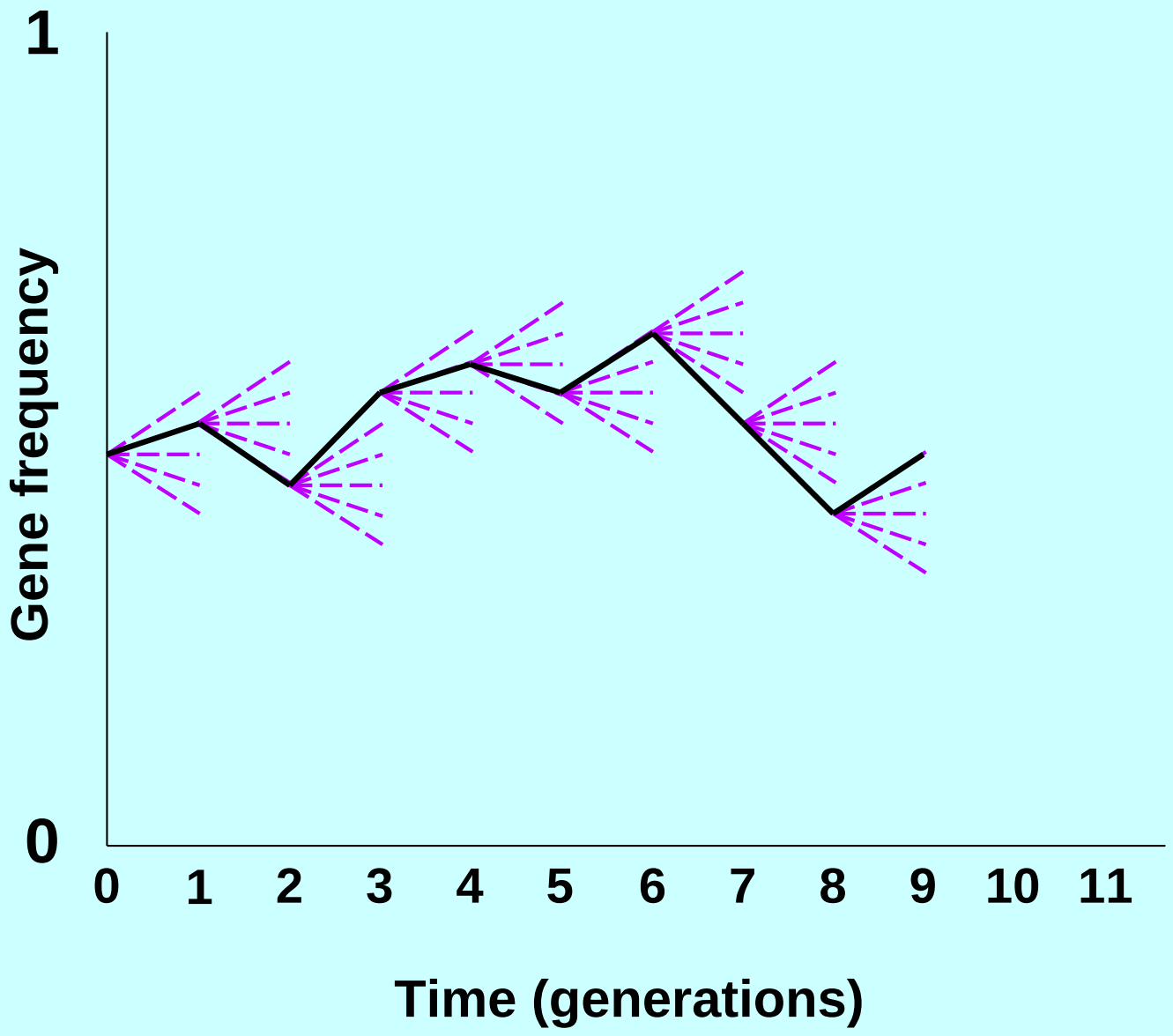
Genetic drift



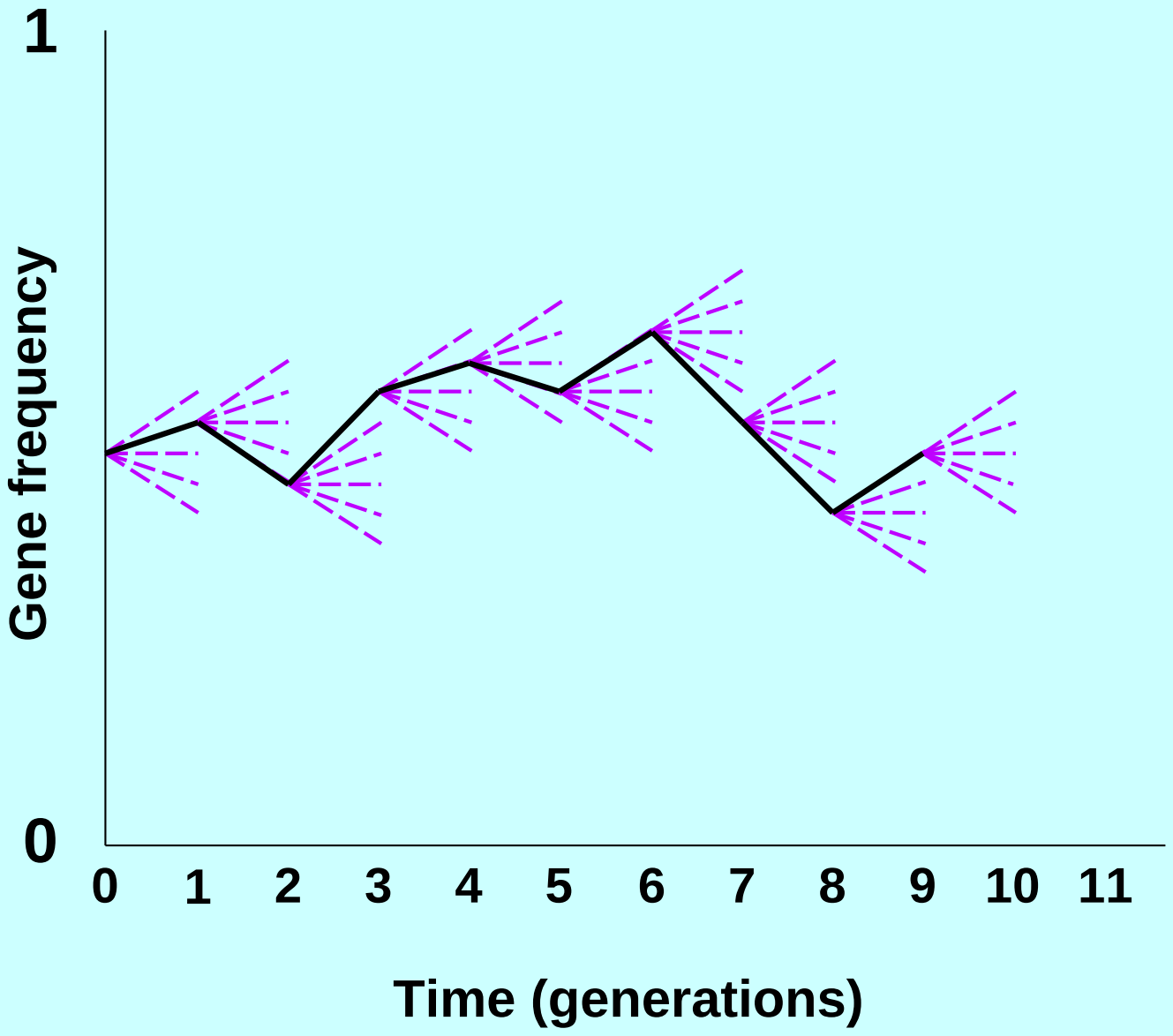
Genetic drift



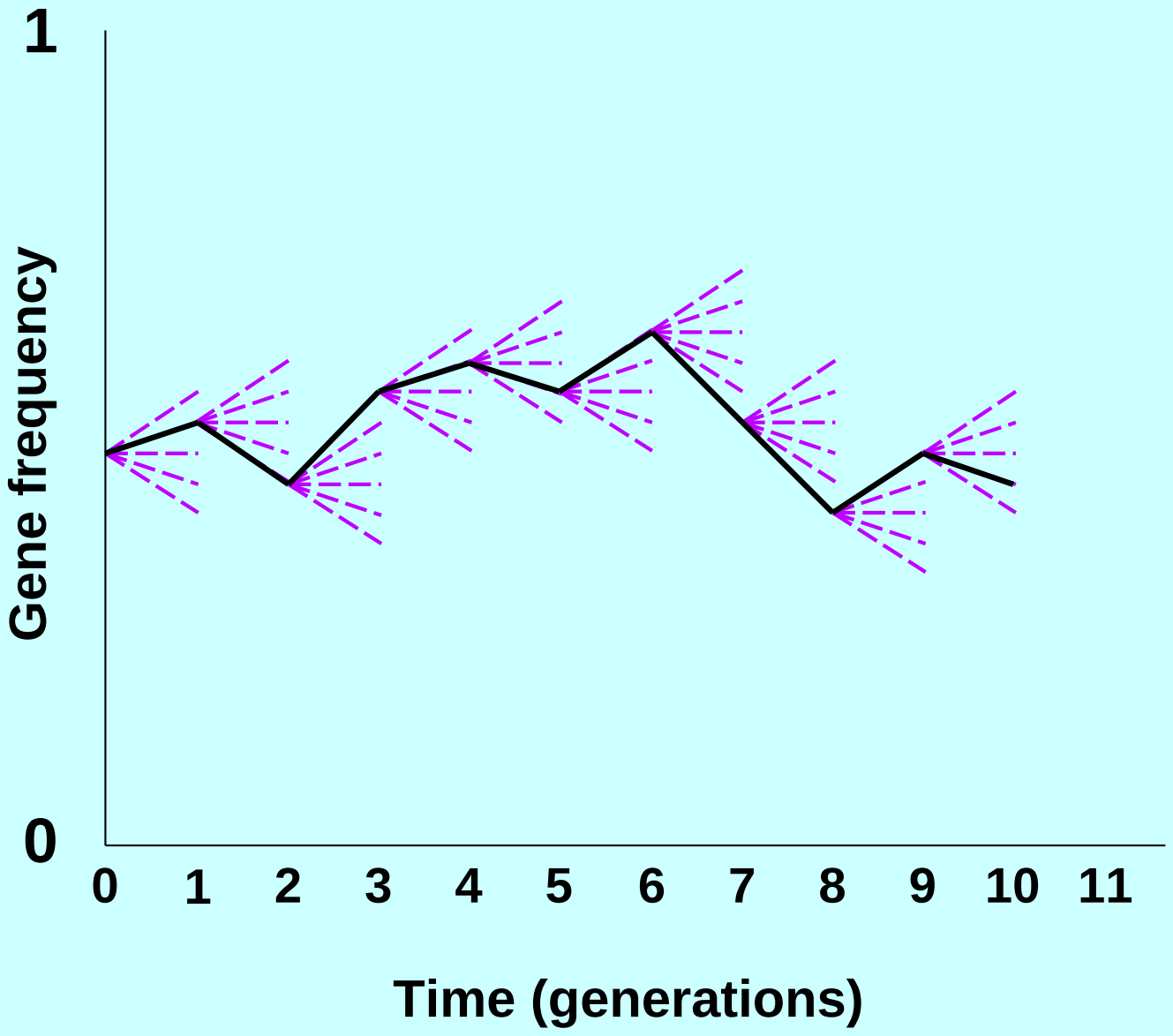
Genetic drift



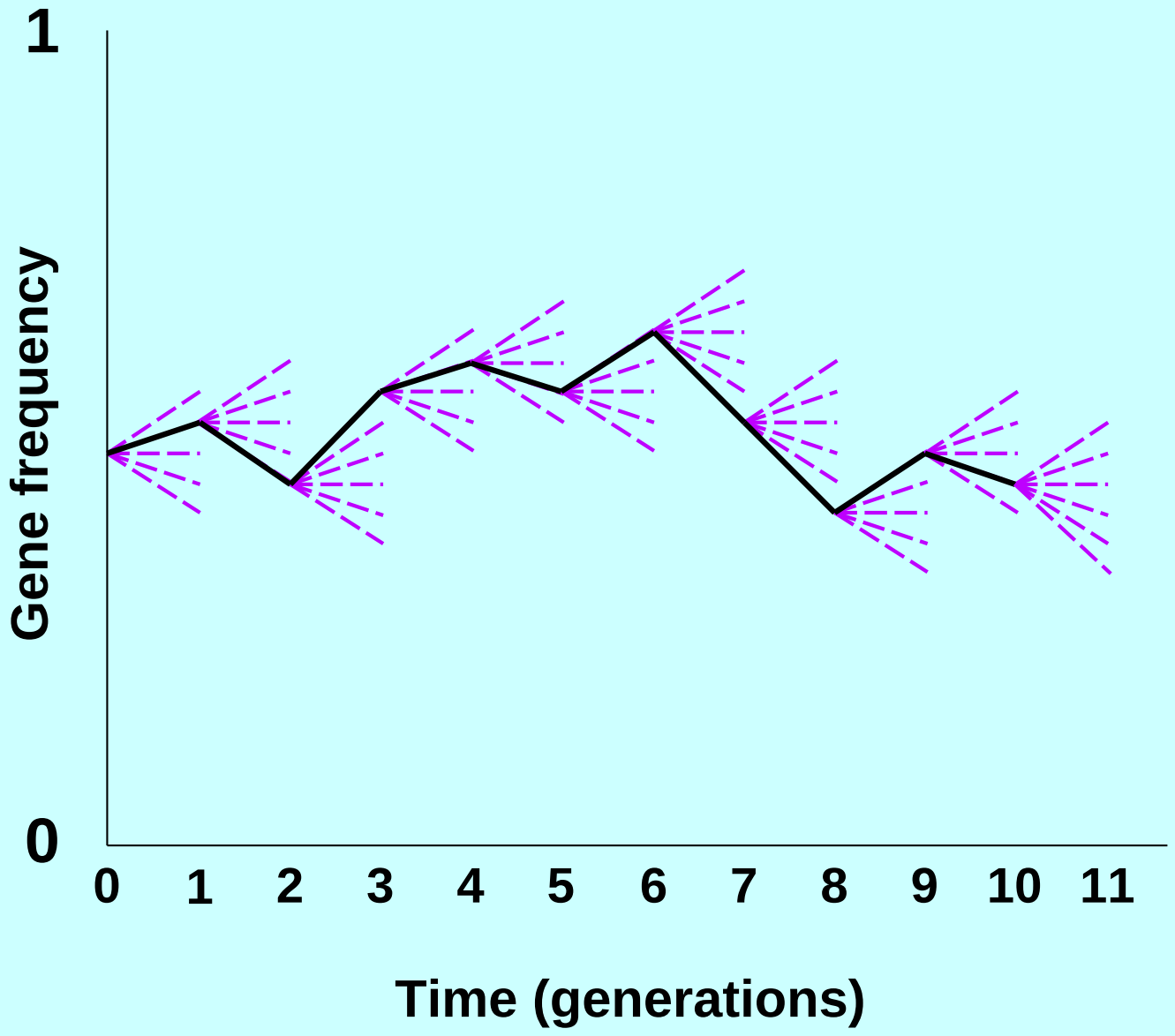
Genetic drift



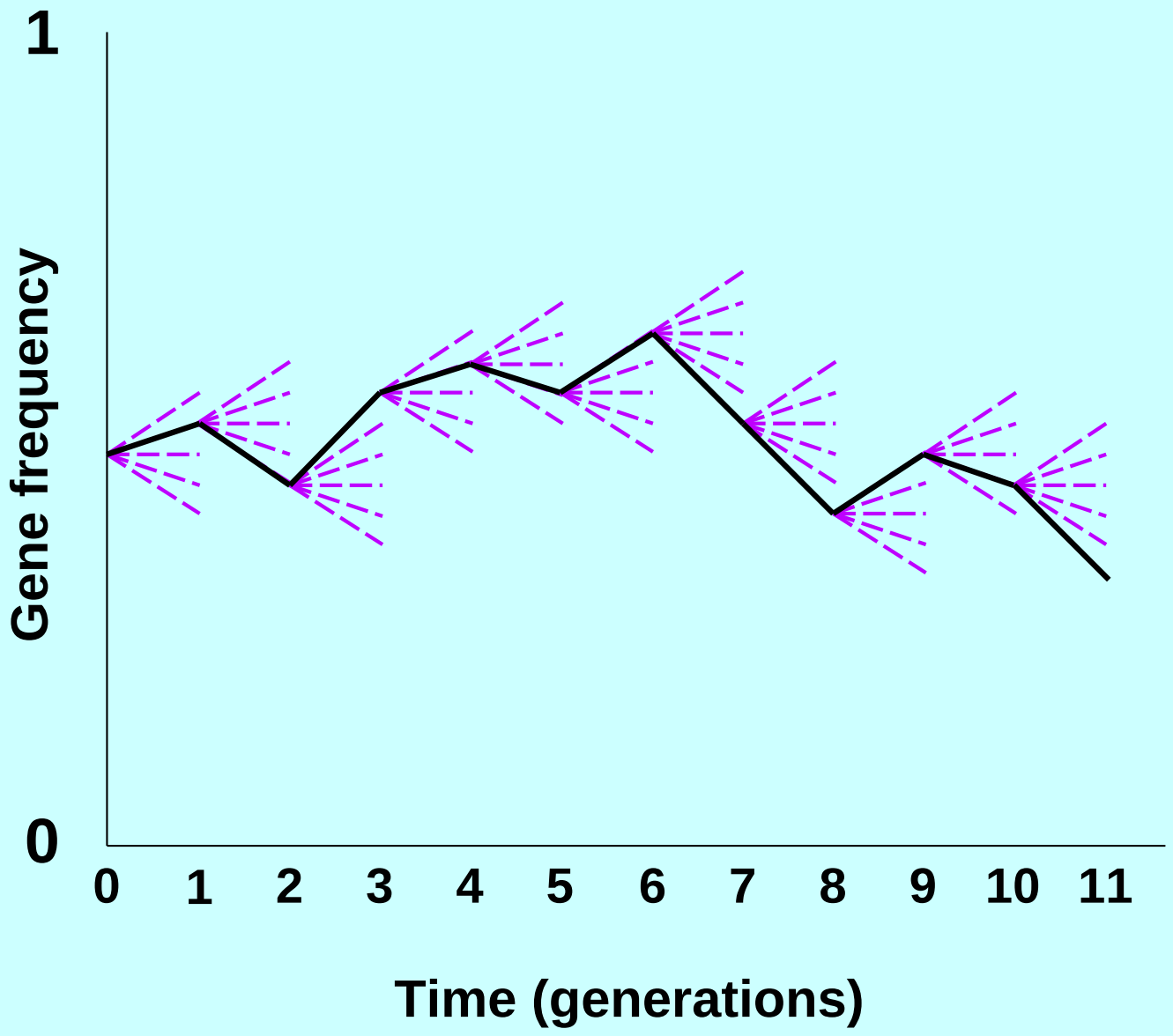
Genetic drift



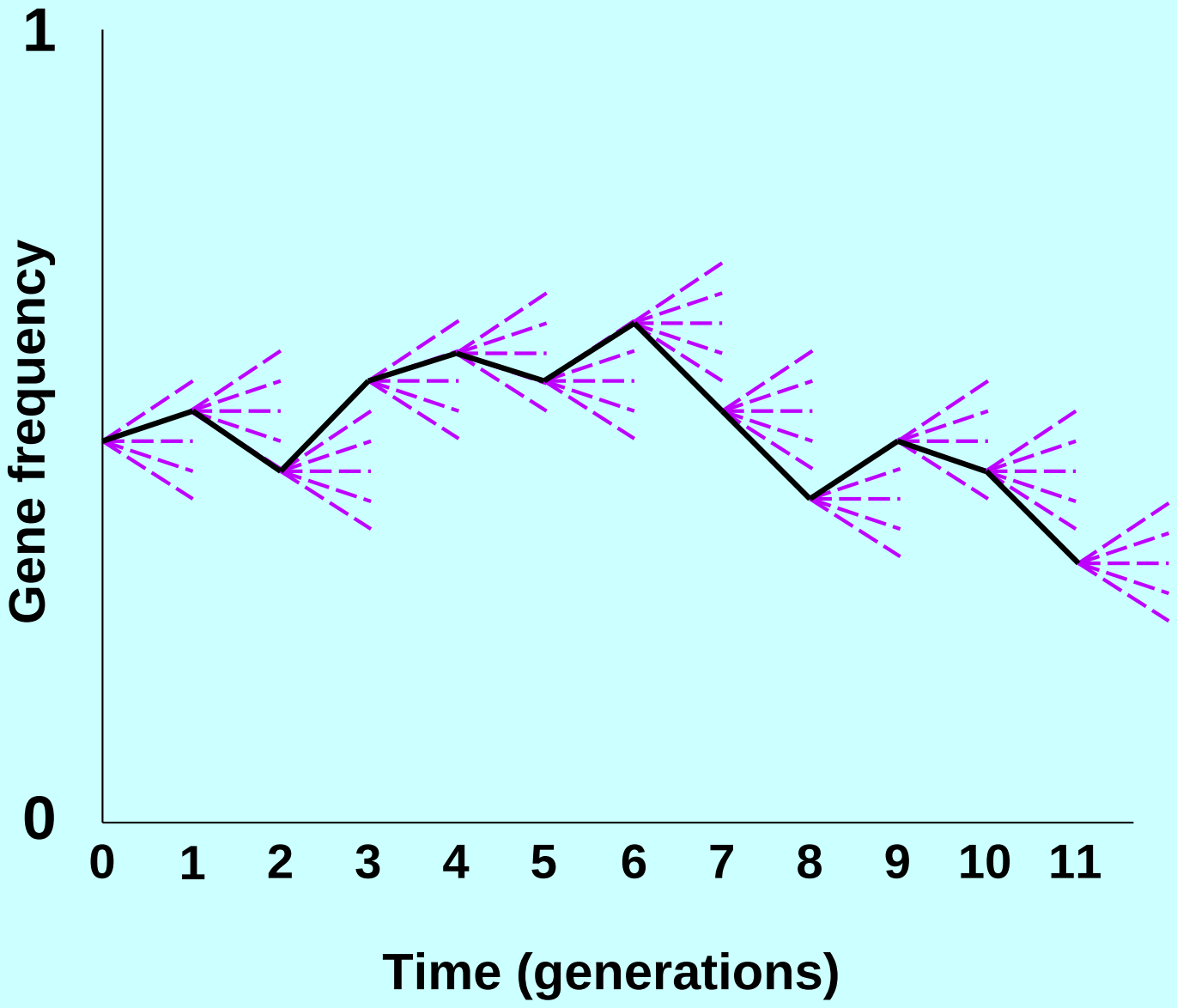
Genetic drift



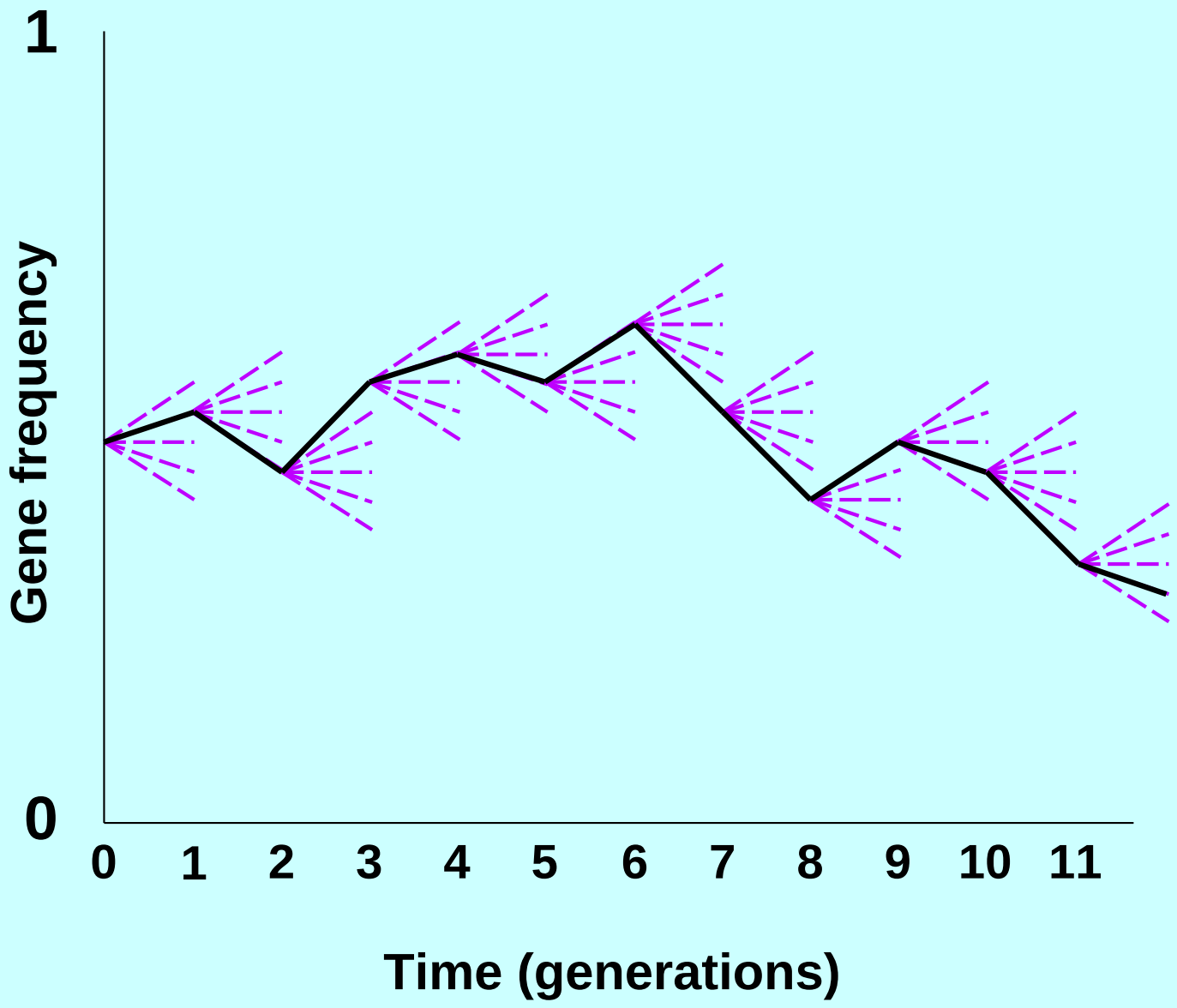
Genetic drift



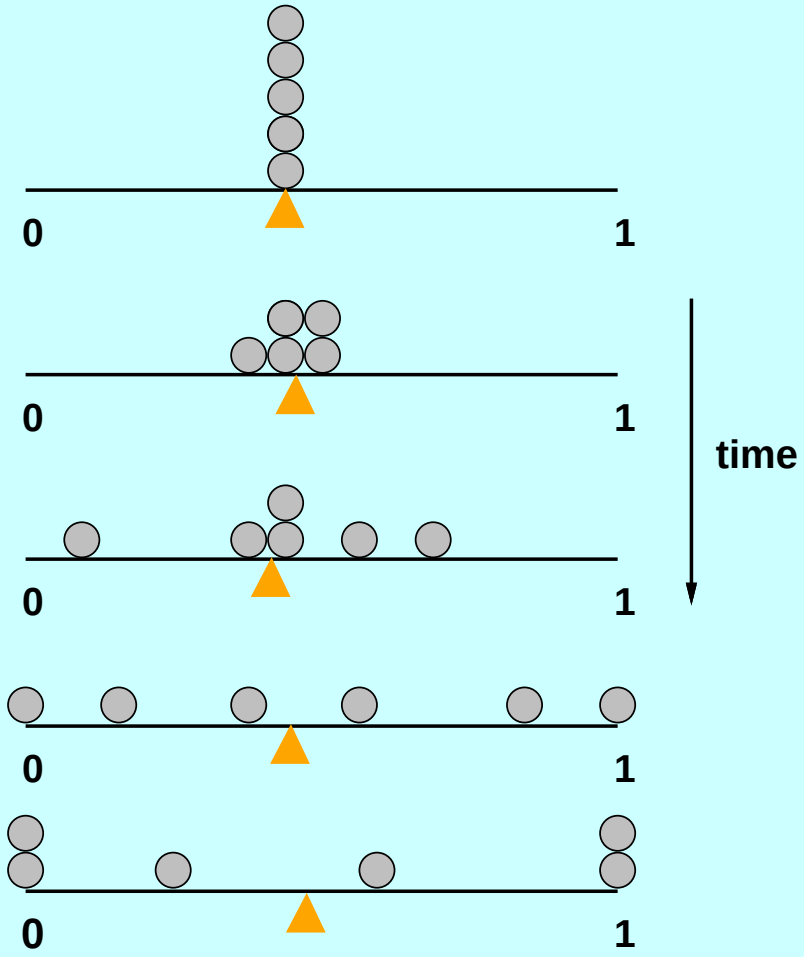
Genetic drift



Genetic drift

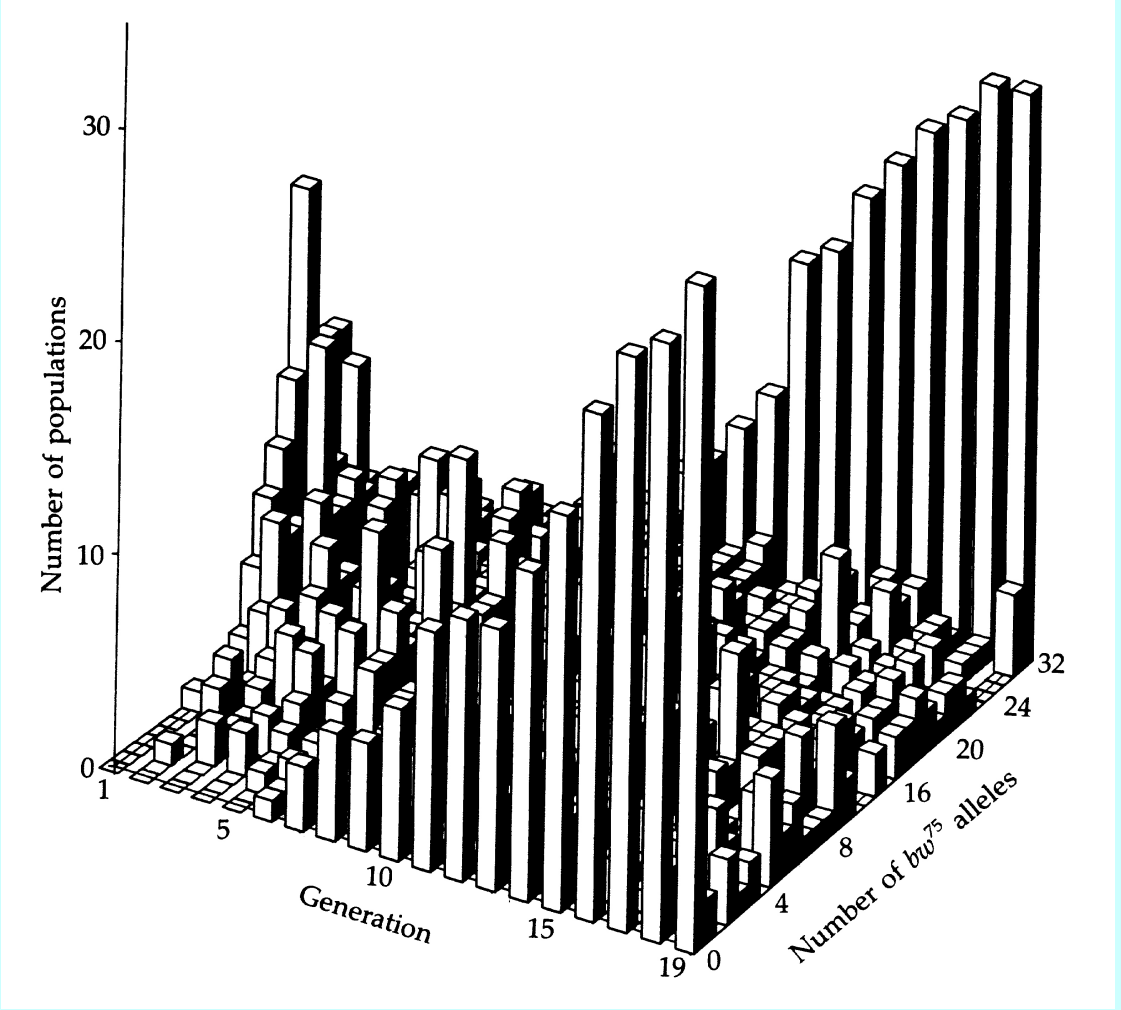


Distribution of gene frequencies with drift



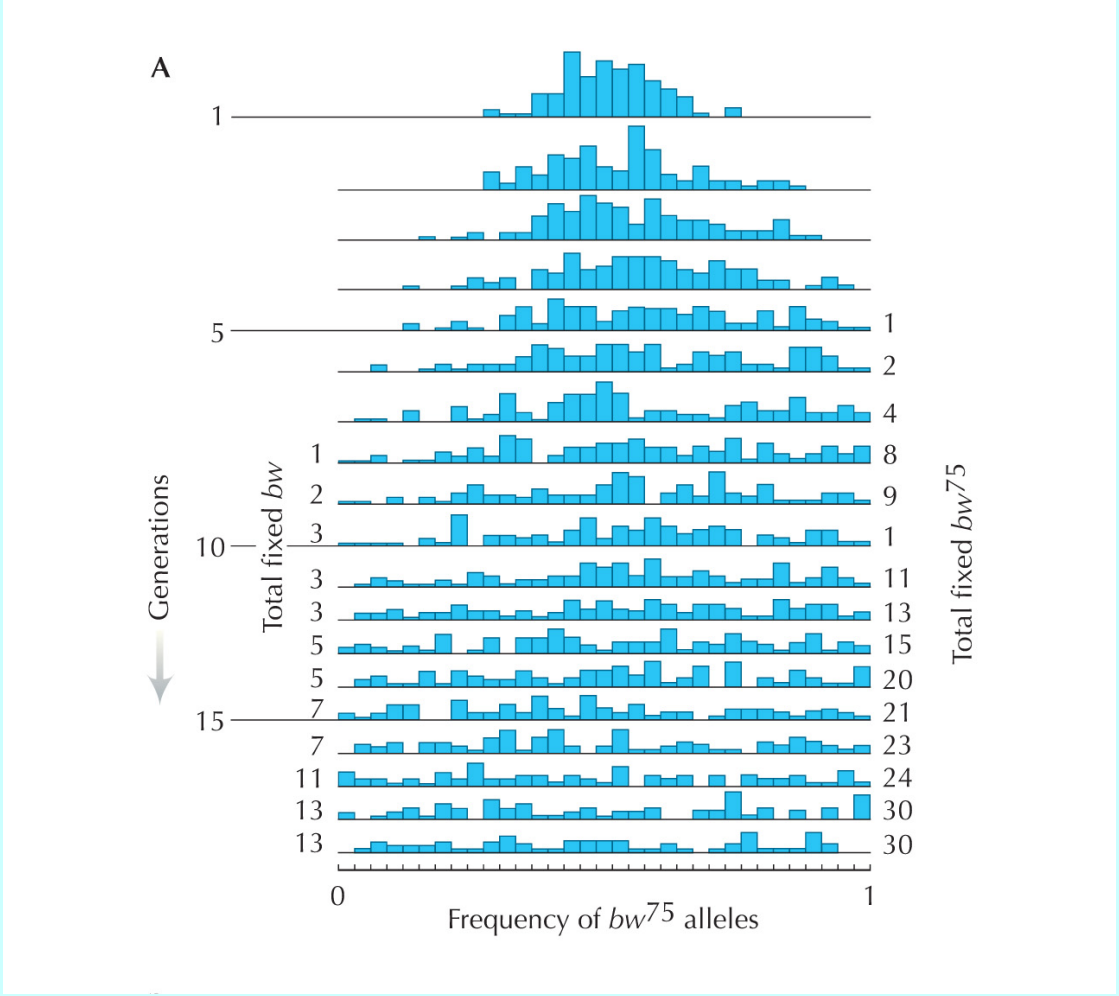
Note that although the individual populations wander, their average hardly moves (not at all when we have infinitely many populations)

Genetic drift in some small populations



from Hartl and Clark,
Principles of Population Genetics

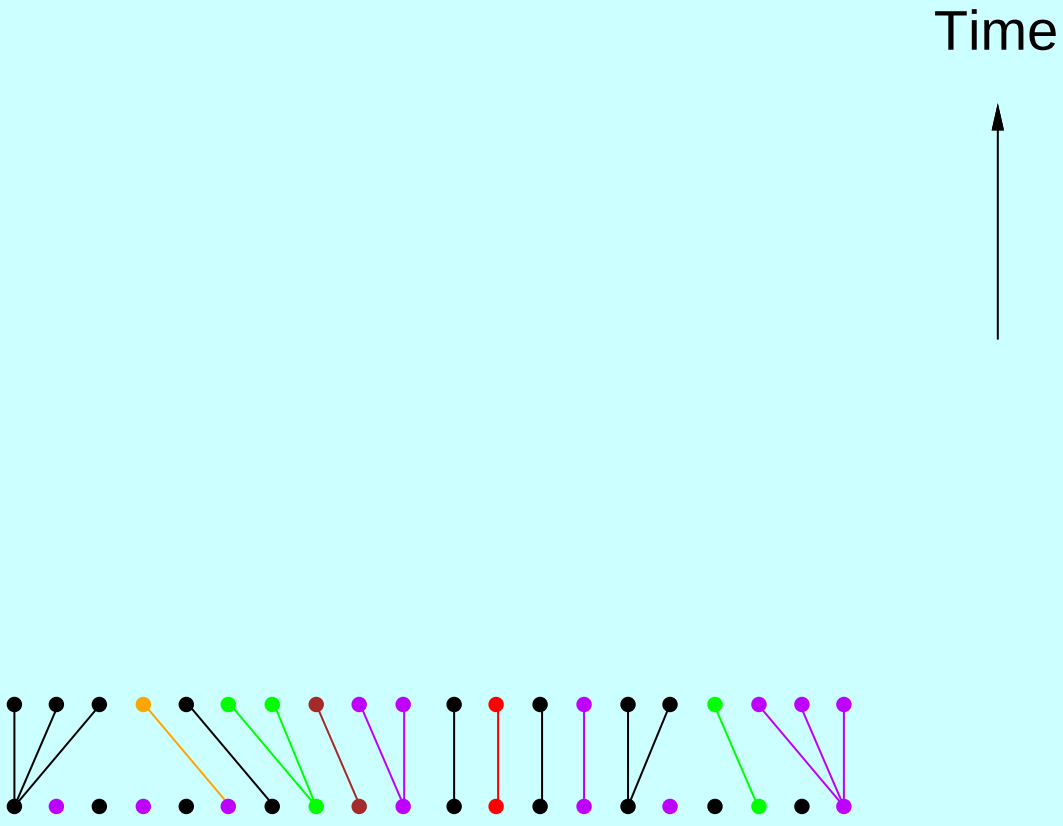
They’re real (lab) populations of *Drosophila*



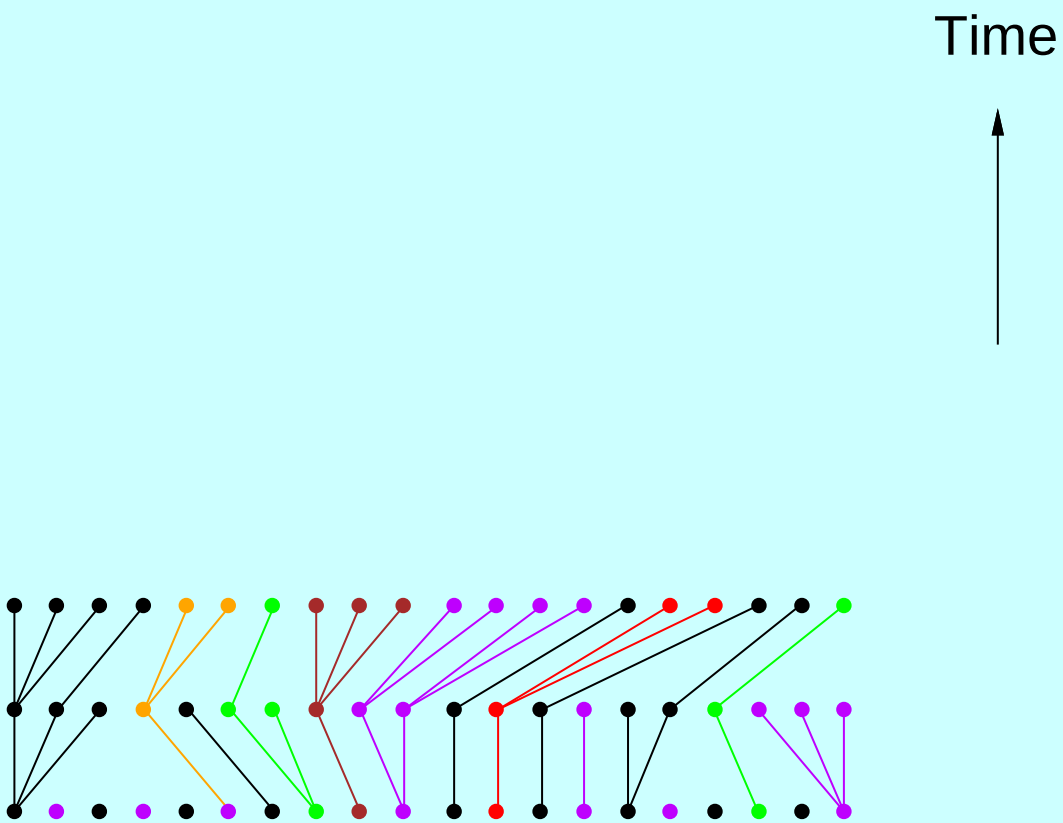
from Barton et al., *Evolution*

Peter Buri. 1956. Gene frequency in small populations of mutant Drosophila. *Evolution* **10 (4):** 367-402.

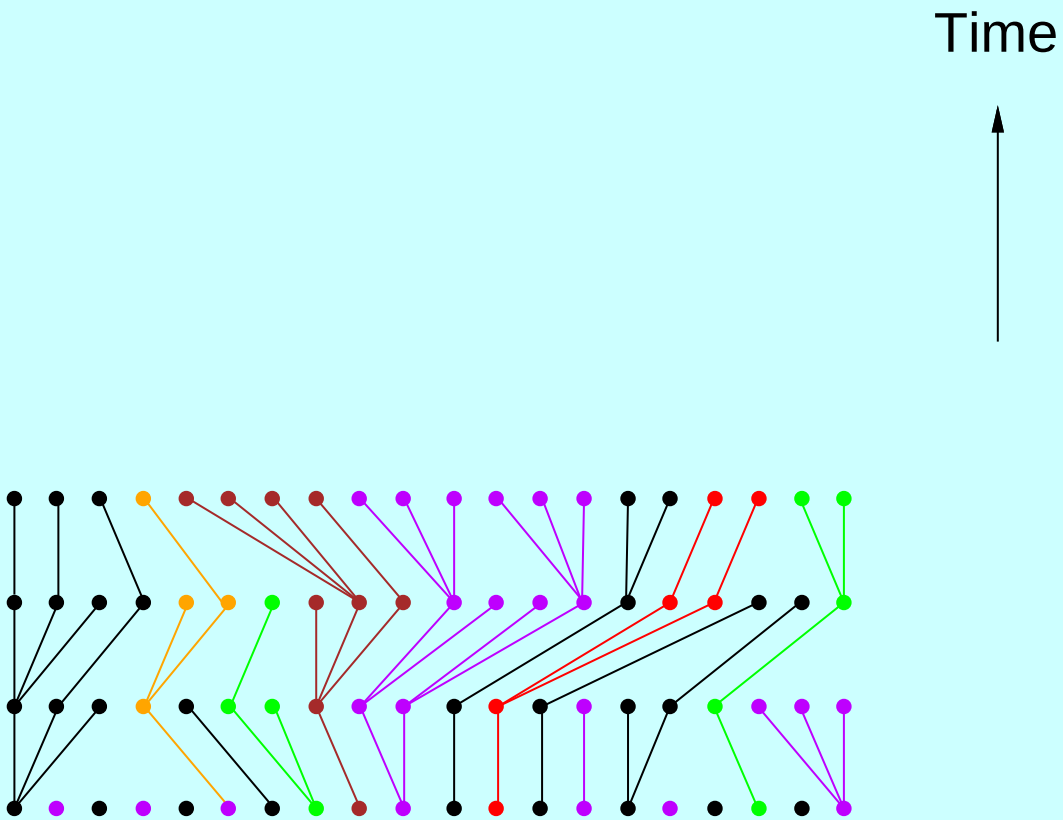
Some copies happen to have more descendants



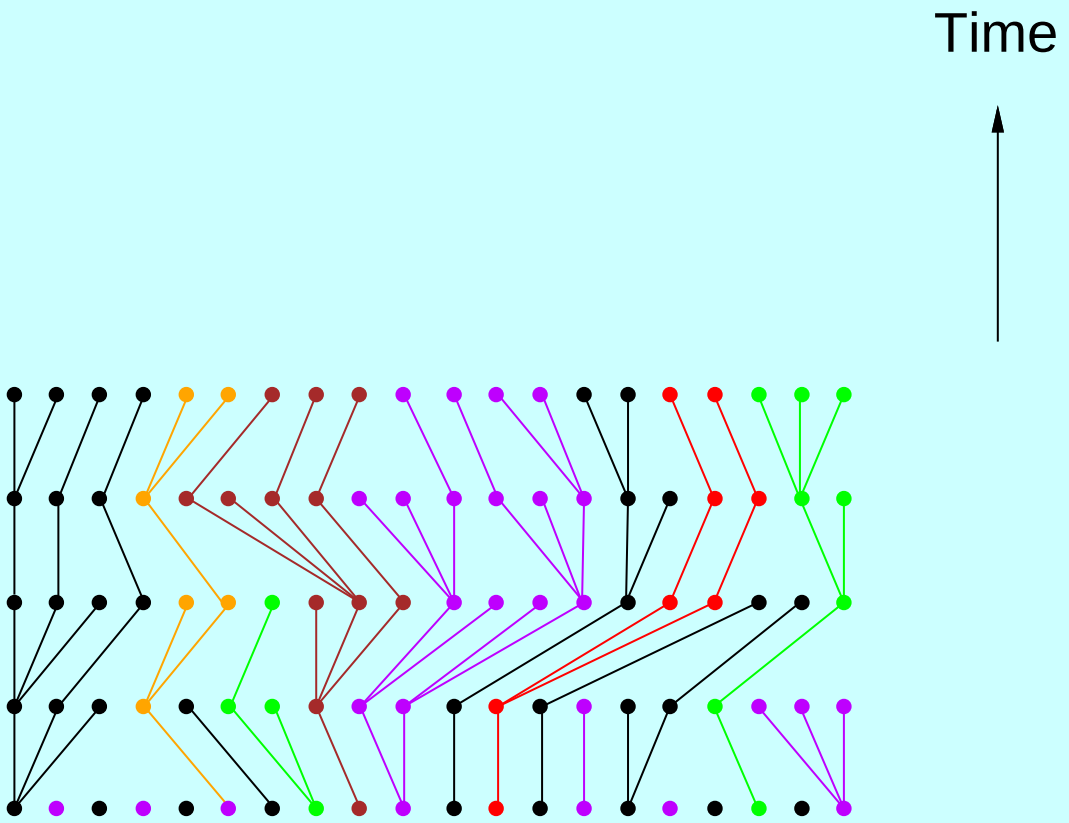
Some copies happen to have more descendants



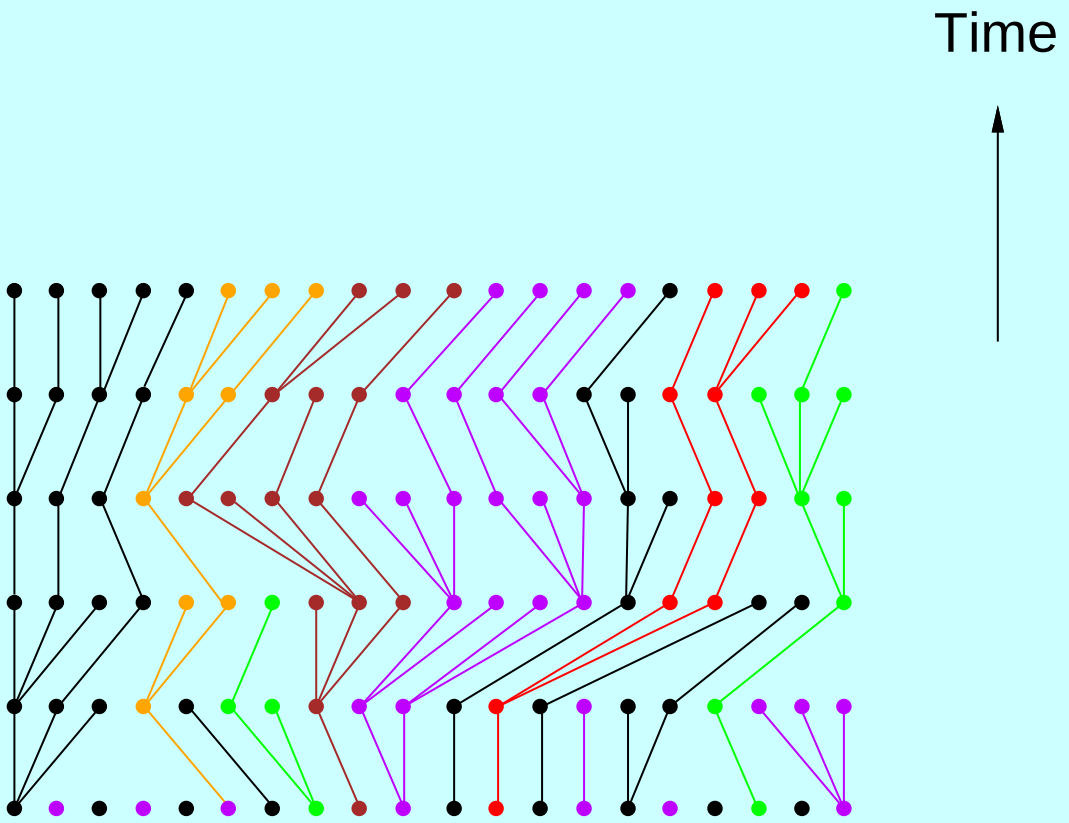
Some copies happen to have more descendants



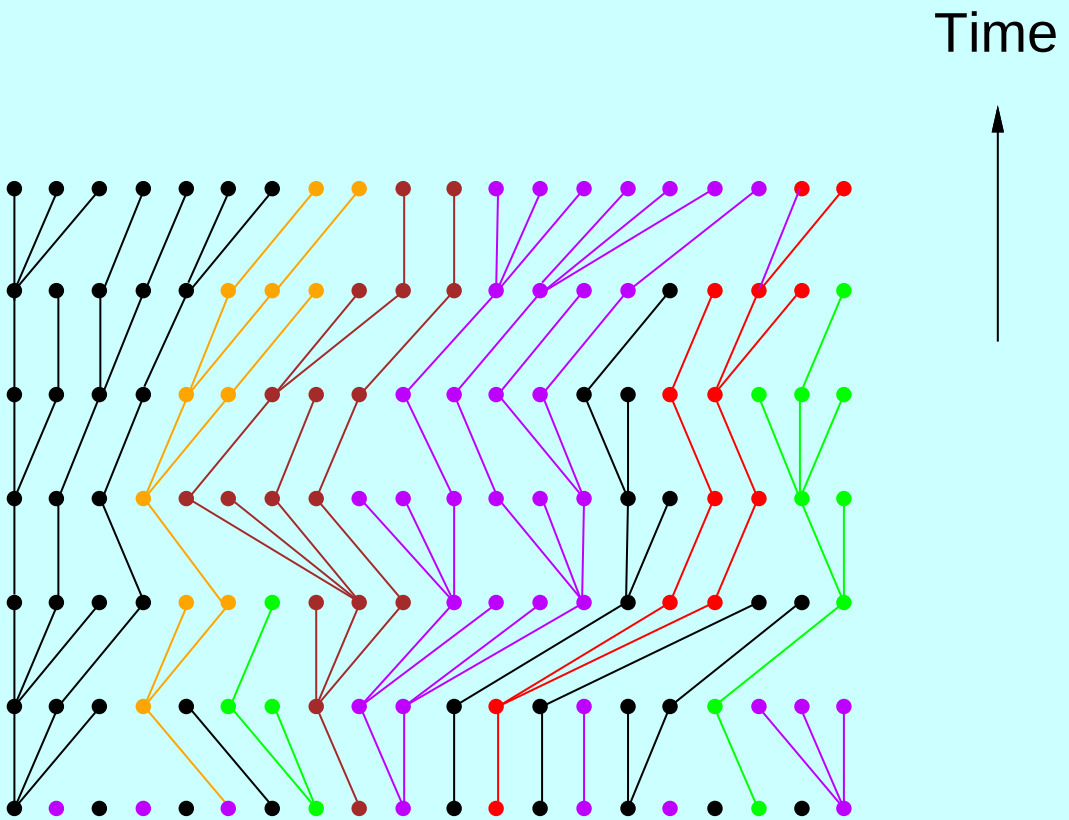
Some copies happen to have more descendants



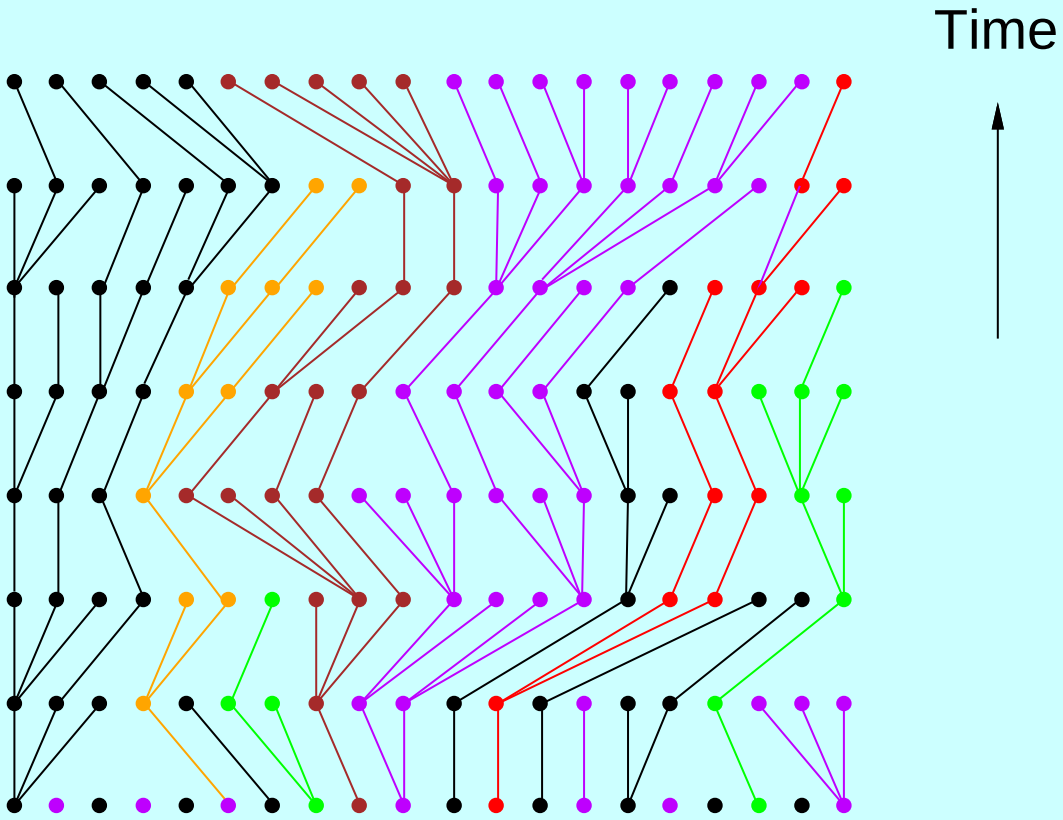
Some copies happen to have more descendants



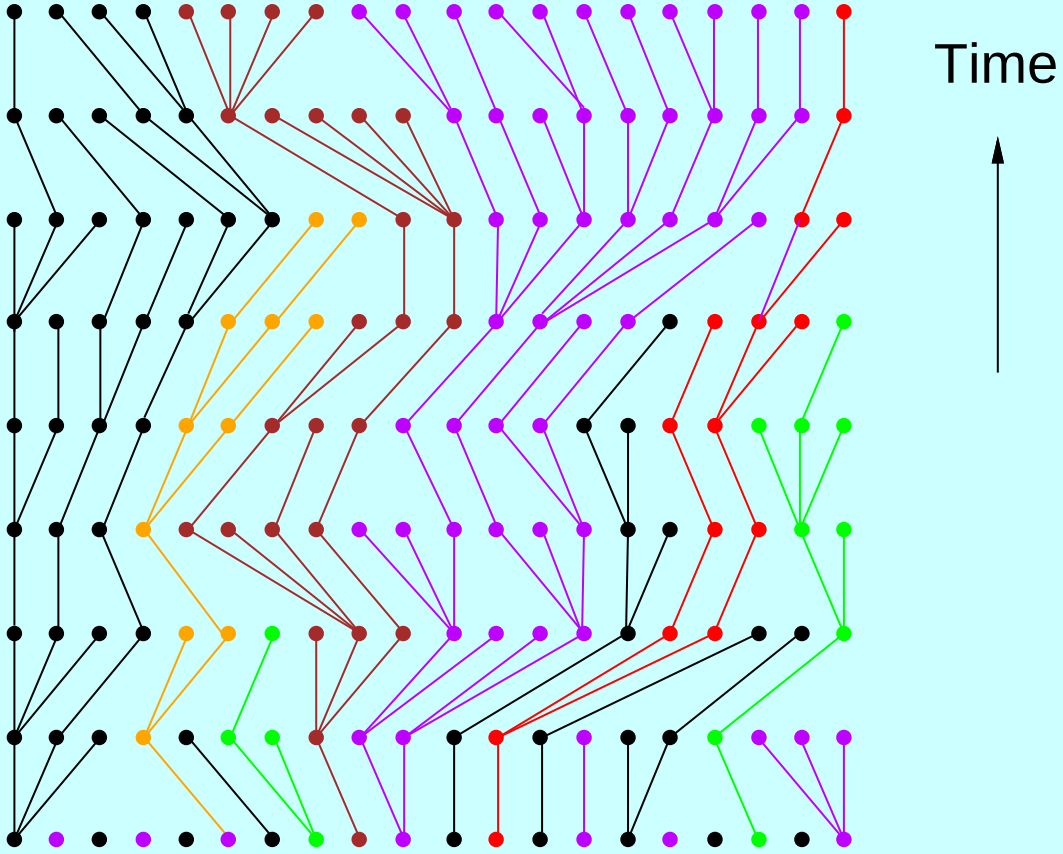
Some copies happen to have more descendants



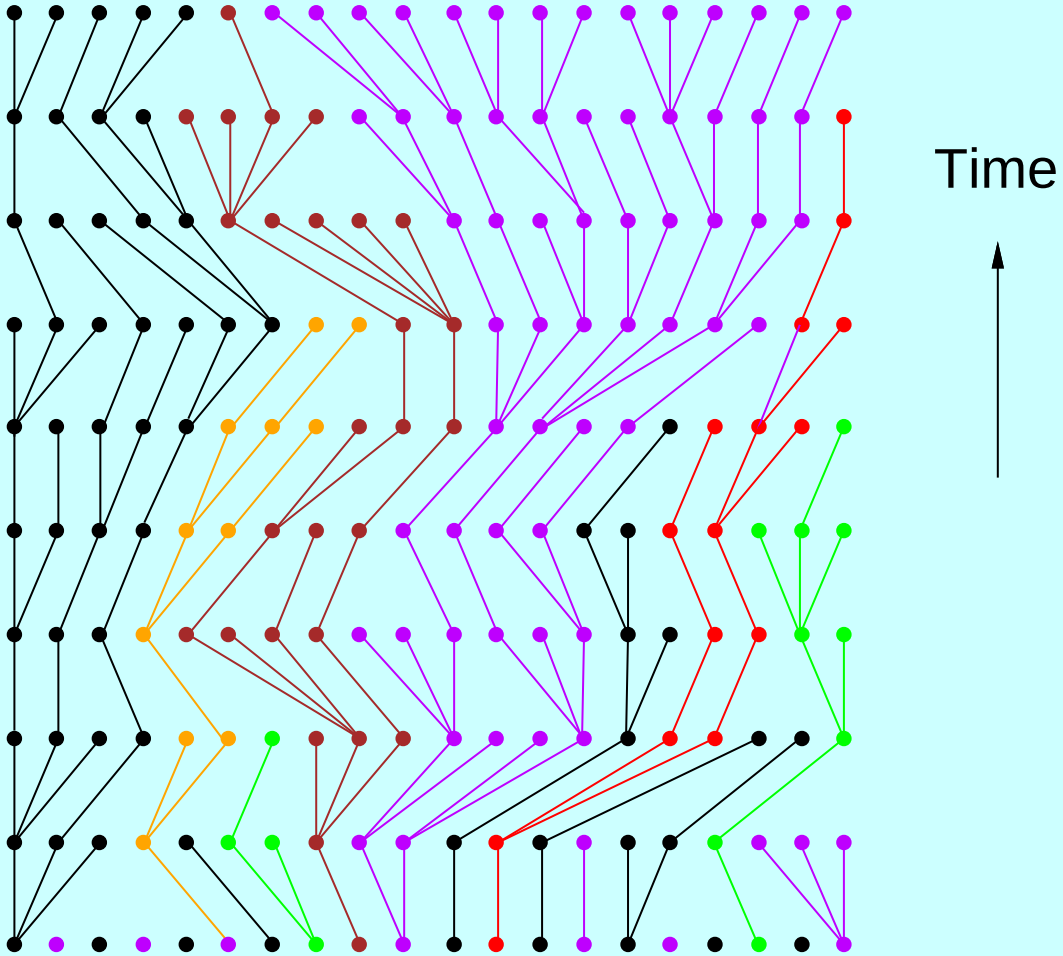
Some copies happen to have more descendants



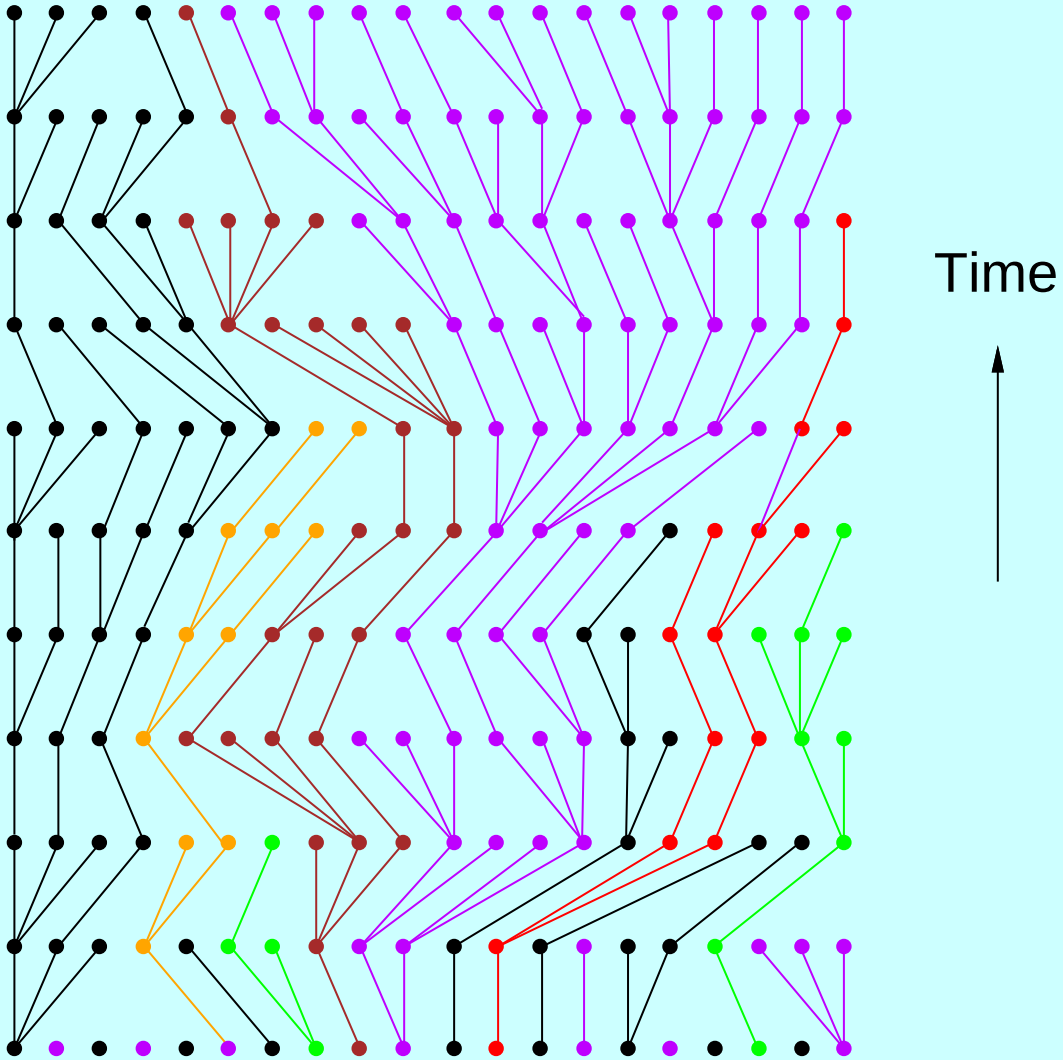
Some copies happen to have more descendants



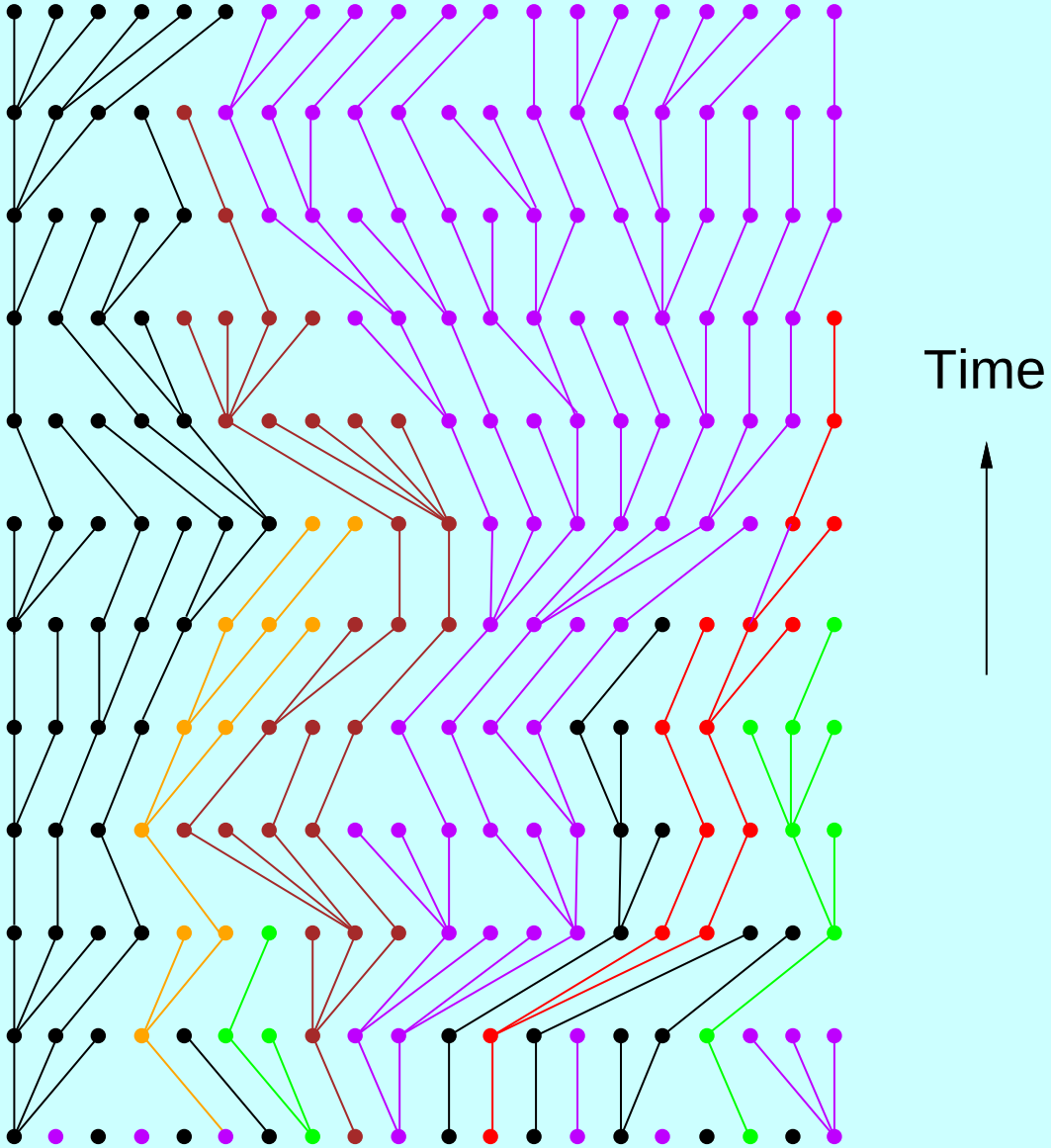
Some copies happen to have more descendants



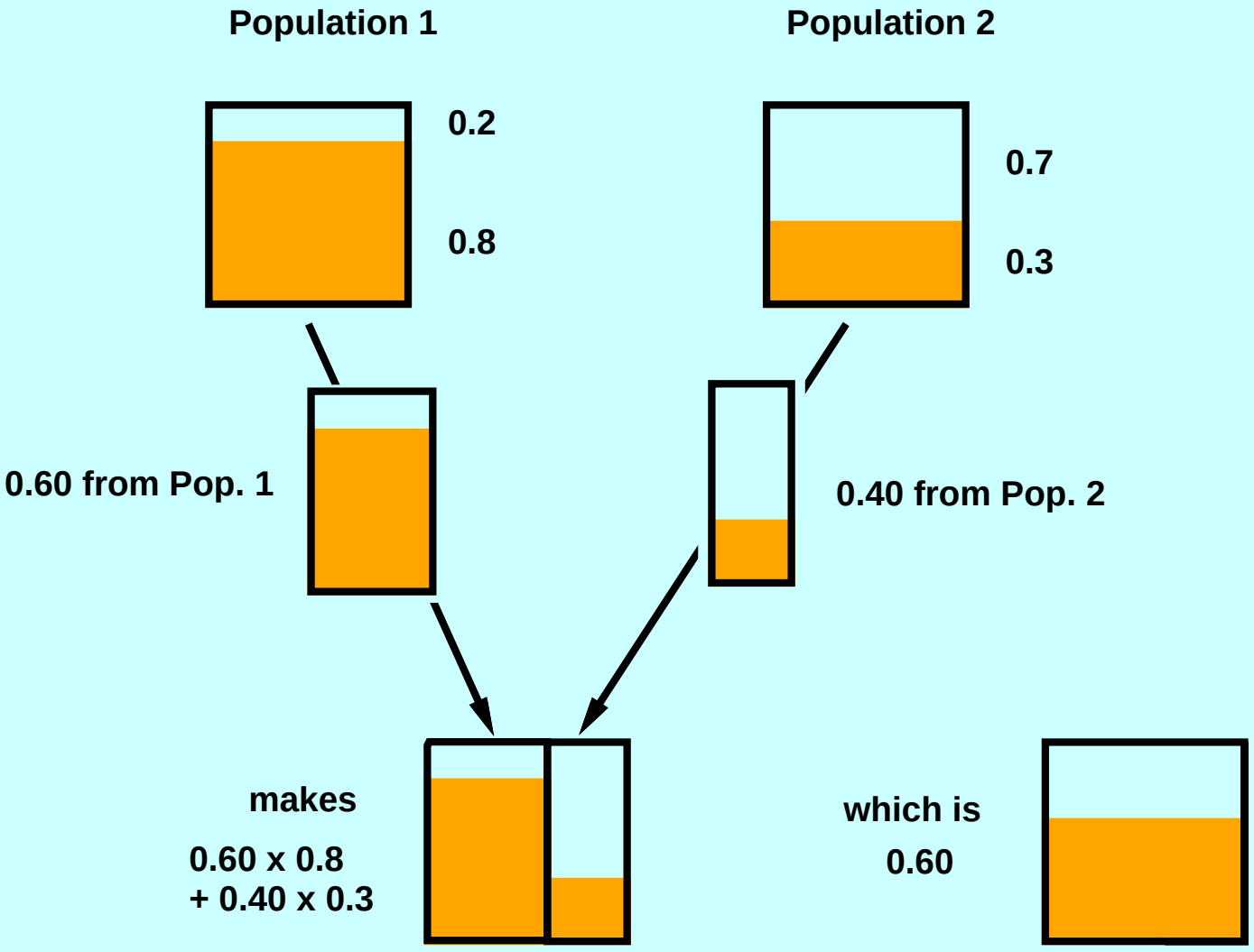
Some copies happen to have more descendants



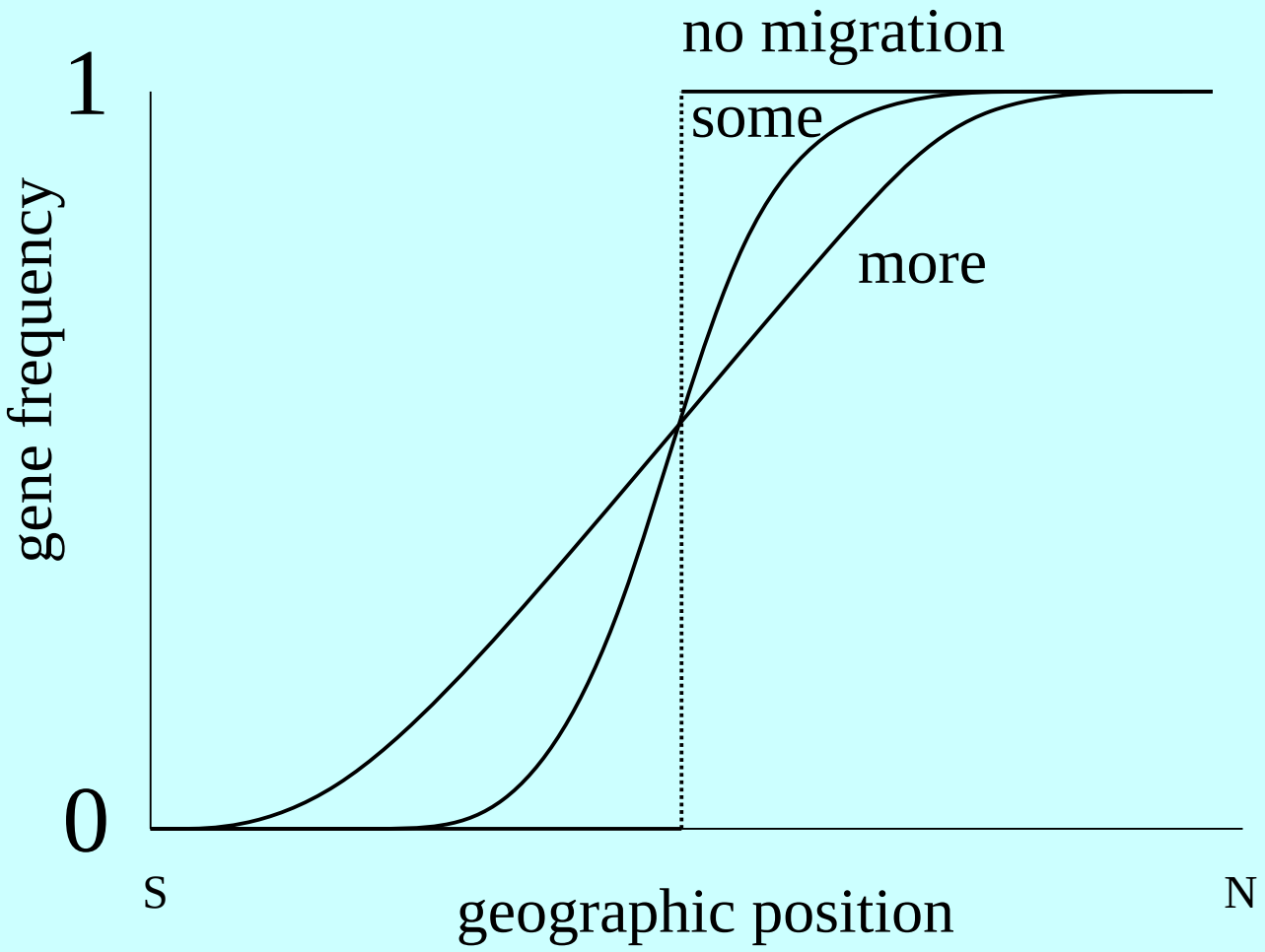
Some copies happen to have more descendants



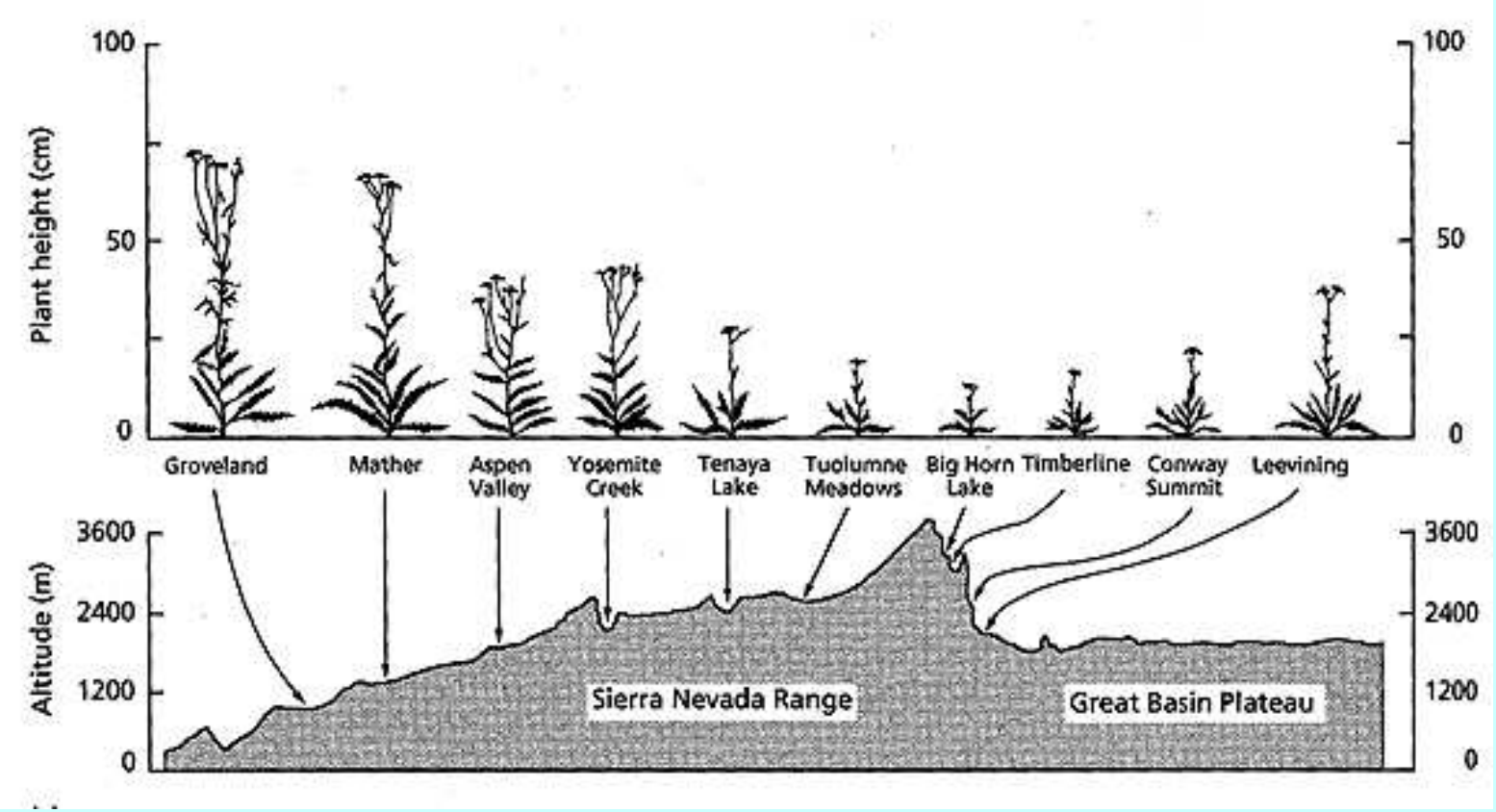
Averaging of gene frequencies when populations admix



A cline (name by Julian Huxley)



A famous common-garden experiment



Clausen, Keck and Hiesey's (1949) common-garden experiment in *Achillea lanulosa*

Heavy metal

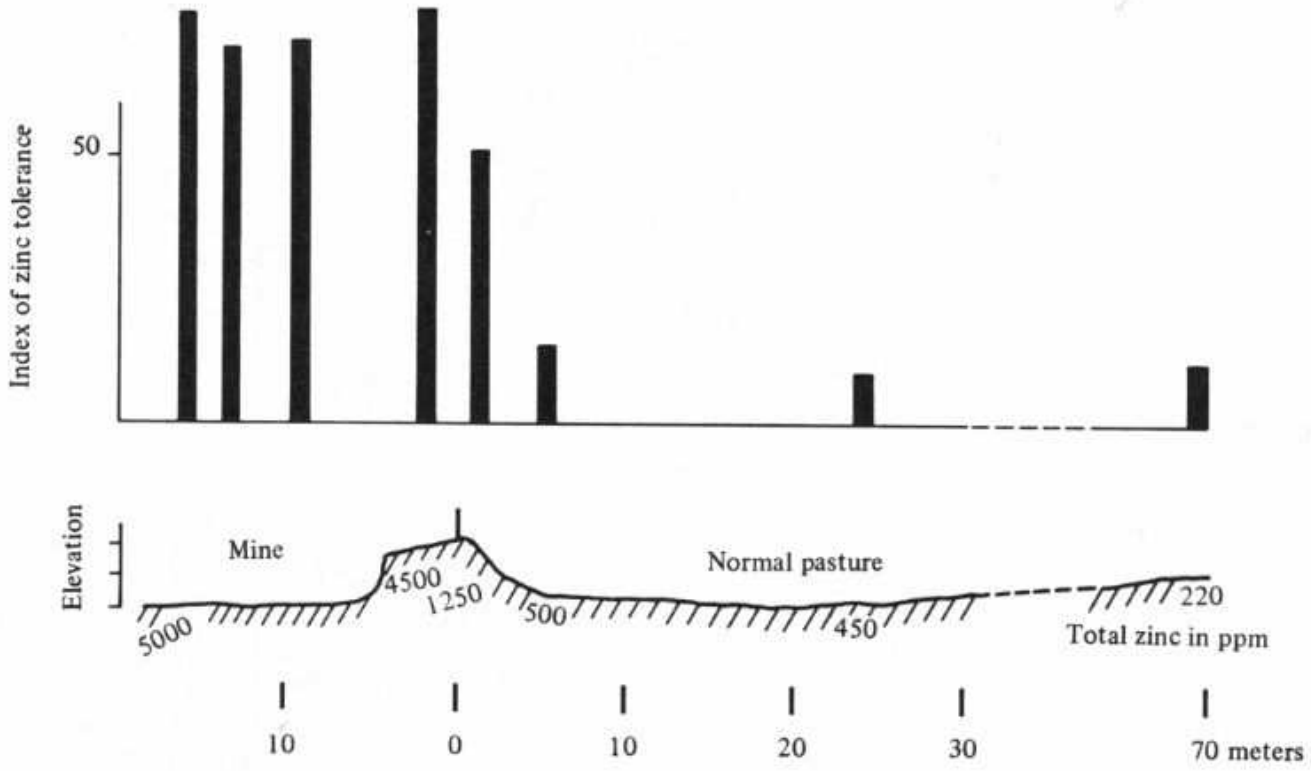


FIGURE 0.5. *The evolution of zinc resistance in grasses over a very fine spatial scale. The top graph illustrates the degree of zinc tolerance exhibited by plants collected from several places along a transect of approximately* 100 *meters in length. The lower graph illustrates the amount of zinc in the soil along the transect. Note the abrupt drop in zinc concentration at the boundary between the mine and the pasture.* [From S. K. Jain and A. D. Bradshaw (1966), Evolutionary divergence among adjacent plant populations I, *Heredity* **21**: 407–441.]

House sparrows

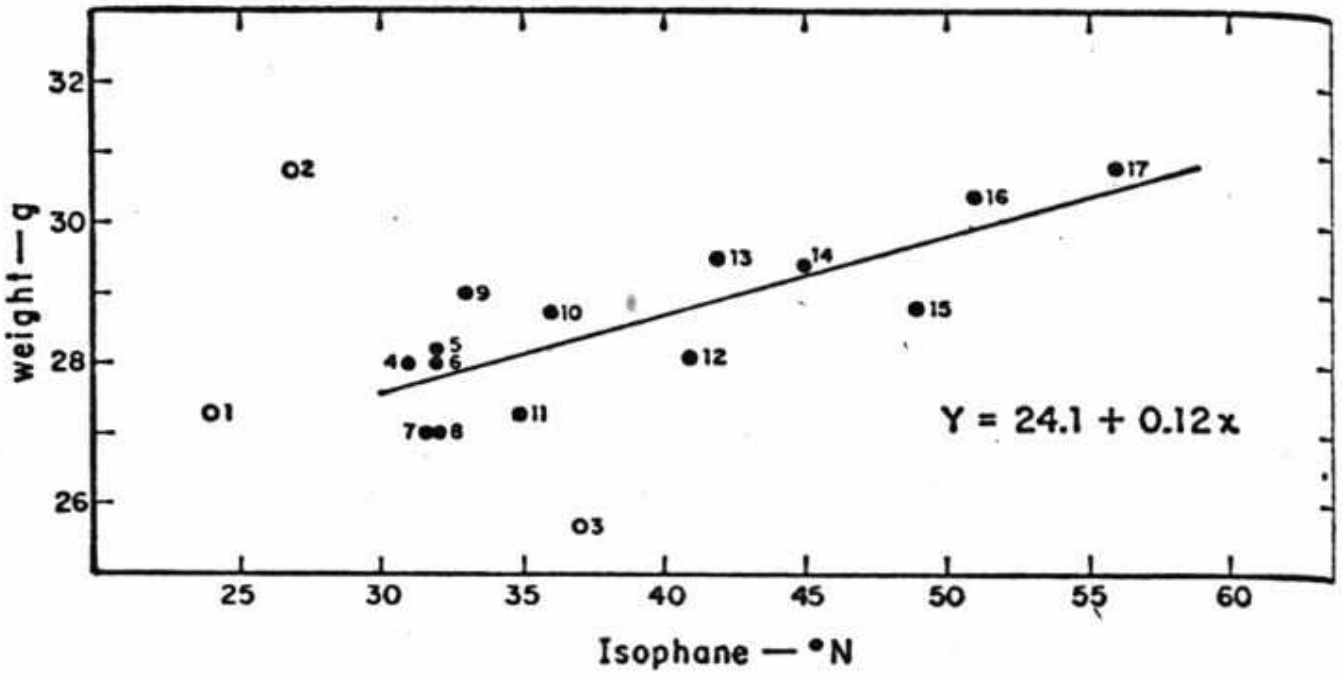


FIGURE 5. Mean body weights of adult male house sparrows plotted against isophanes (see text for explanation). Localities: 1, Oaxaca City, Mexico; 2, Progreso, Tex.; 3, Mexico City, Mexico; 4, Houston, Tex.; 5, Los Angeles, Calif.; 6, Austin, Tex.; 7, Death Valley, Calif.; 8, Phoenix, Ariz.; 9, Baton Rouge, La.; 10, Sacramento, Calif.; 11, Oakland, Calif.; 12, Las Cruces, N.M.; 13, Lawrence, Kan.; 14, Vancouver, B.C.; 15, Salt Lake City, Utah; 16, Montreal, Quebec; 17, Edmonton, Alberta. The regression line is based on data from localities 4 to 17.

House sparrows

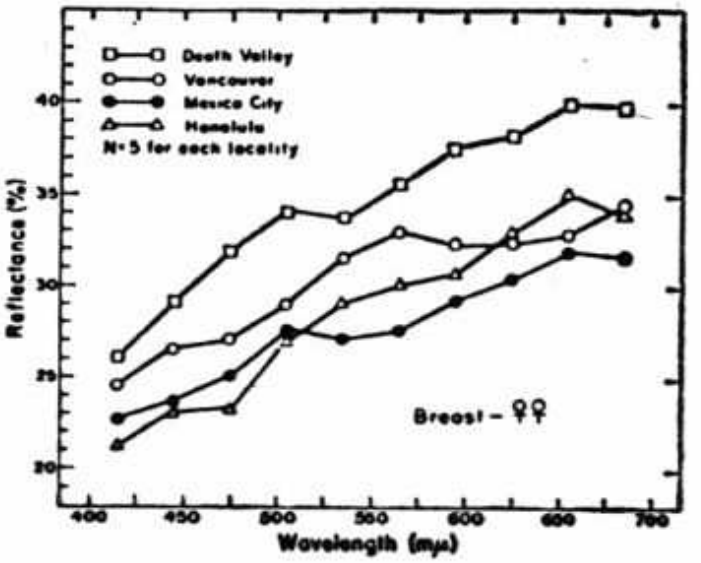


FIGURE 2. Spectral reflectance curves for the breast of female house sparrows from Honolulu, Hawaii, and several North American localities.

Mutation Rates

Coat color mutants in mice. From

Schlager G. and M. M. Dickie. 1967. Spontaneous mutations and mutation rates in the house mouse. *Genetics* **57:** 319-330

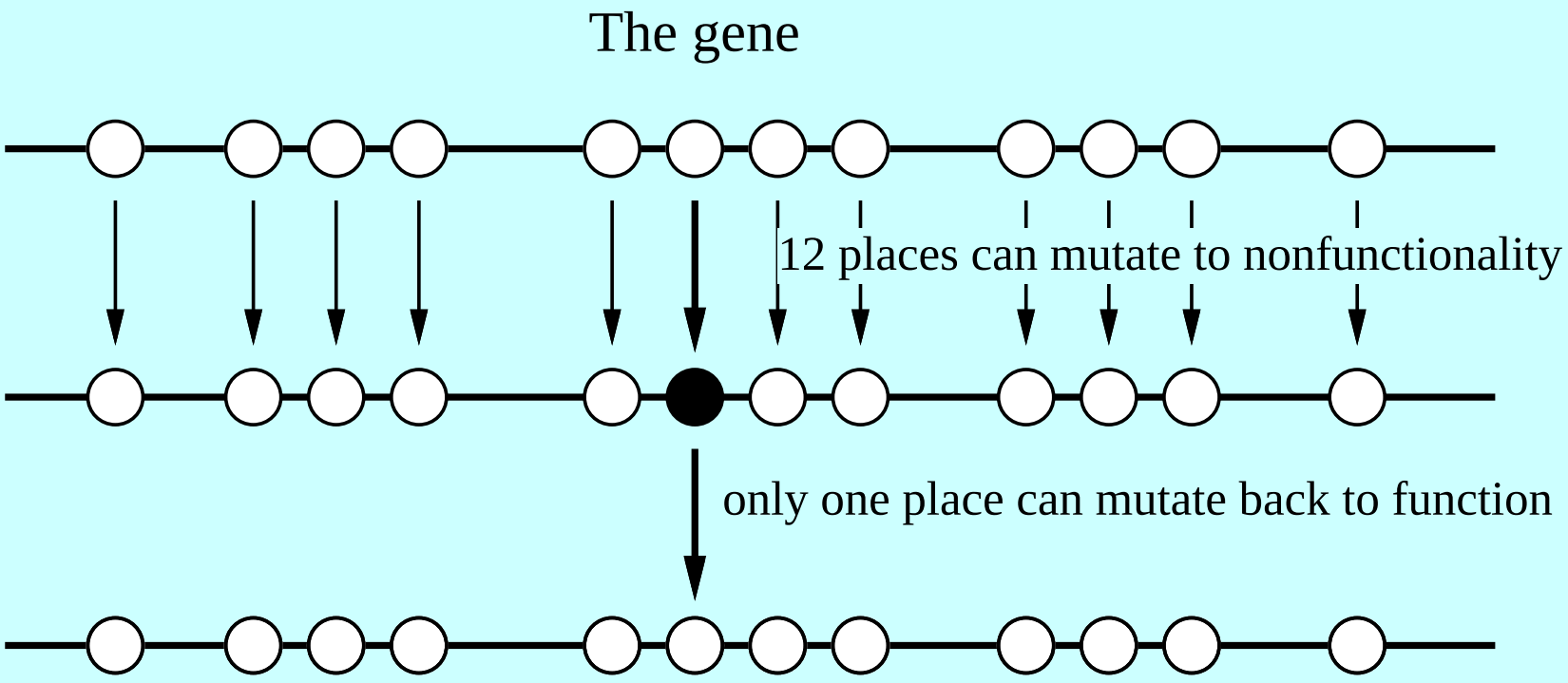
Locus	Gametes tested	No. of Mutations	Rate
Nonagouti	67,395	3	4.4×10^{-6}
Brown	919,619	3	3.3×10^{-6}
Albino	150,391	5	33.2×10^{-6}
Dilute	839,447	10	11.9×10^{-6}
Leaden	243,444	4	16.4×10^{-6}
Total	2,220,376	25	11.2×10^{-6}

Mutation rates in humans

Trait	Population examined	Mutation rate	Number of mutants/10^6 gametes	Authors
A. Autosomal mutations				
Achondroplasia	Denmark	1×10^{-5}	10	Mørch, corrected by Slatis
	Northern Ireland	1.3×10^{-5}	13	Stevenson
	Germany (Reg. Bez. Münster)	$6–9 \times 10^{-6}$	6–9	Schiemann
Aniridia	Denmark	$2.9(-5) \times 10^{-6}$	2.9(–5)	Møllenbach, corrected by Penrose
	Michigan (U.S.A.)	2.6×10^{-6}	2.6	Shaw et al.
Dystrophia myotonica	Northern Ireland	8×10^{-6}	8	Lynas
	Switzerland	1.1×10^{-5}	11	Klein, corrected by Todorov et al.
Retinoblastoma	England, Michigan (U.S.A.), Switzerland, Germany	$6–7 \times 10^{-6}$	6–7	Vogel
	Hungary	6×10^{-6}	6	Czeizel et al.
	The Netherlands	1.23×10^{-5}	12.3	Schappert-Kimmijser et al.
	Japan	8×10^{-6}	8	Matsunaga
	France	5×10^{-6}	5	Briart-Guillemot et al.
Acrocephalosyndactyly (Apert's syndrome)	England	3×10^{-6}	3	Blank
	Germany (Reg. Bez. Münster)	4×10^{-6}	4	Tünte and Lenz
Osteogenesis imperfecta	Sweden	$0.7–1.3 \times 10^{-5}$	7–13	Smårs
	Germany (Reg. Bez. Münster)	1.0×10^{-5}	10	Schröder
Tuberous sclerosis (epiloia)	Oxford Regional Hospital Board Area (G.B.)	1.05×10^{-5}	10.5	Nevin and Pearce
	Chinese	6×10^{-6}	6	Singer
Neurofibromatosis	Michigan (U.S.A.)	1×10^{-4}	100	Crowe et al.
	Moscow (U.S.S.R.)	$4.4–4.9 \times 10^{-5}$	44–49	Sergeyev
Polyposis of intestines	Michigan (U.S.A.)	1.3×10^{-5}	13	Reed and Neel
Marfan's syndrome	Northern Ireland	$4.2–5.8 \times 10^{-6}$	4.2–5.8	Lynas
Polycystic disease of the kidneys	Denmark	$6.5–12 \times 10^{-5}$	65–120	Dalgaard
Diaphyseal aclasis (multiple exostoses)	Germany (Reg. Bez. Münster)	$6.3–9.1 \times 10^{-6}$	6.3–9.1	Murken
von Hippel-Lindau syndrome	Germany	1.8×10^{-7}	0.18	Burhorn

Forward vs. back mutations

Why mutants inactivating a functional gene will be more frequent than back mutations



function can sometimes be restored by a "second site" mutation, too

A sequence space

For sequences of length 1000, there are 3 $\times$ 1000 = 3000 “neighbors” one step away in sequence space.

But there are 4^{1000} sequences, which is about 10^{602} in all !

No two of them are more than 1000 steps apart. Hard to draw such a space! How do we ever evolve? Wouldn’t it be impossible to find one of the tiny fraction of possible sequences that would be even marginally functional?

The answer seems to be that the sequences are clustered. An example of such clustering is the English language, as illustrated by a word game:

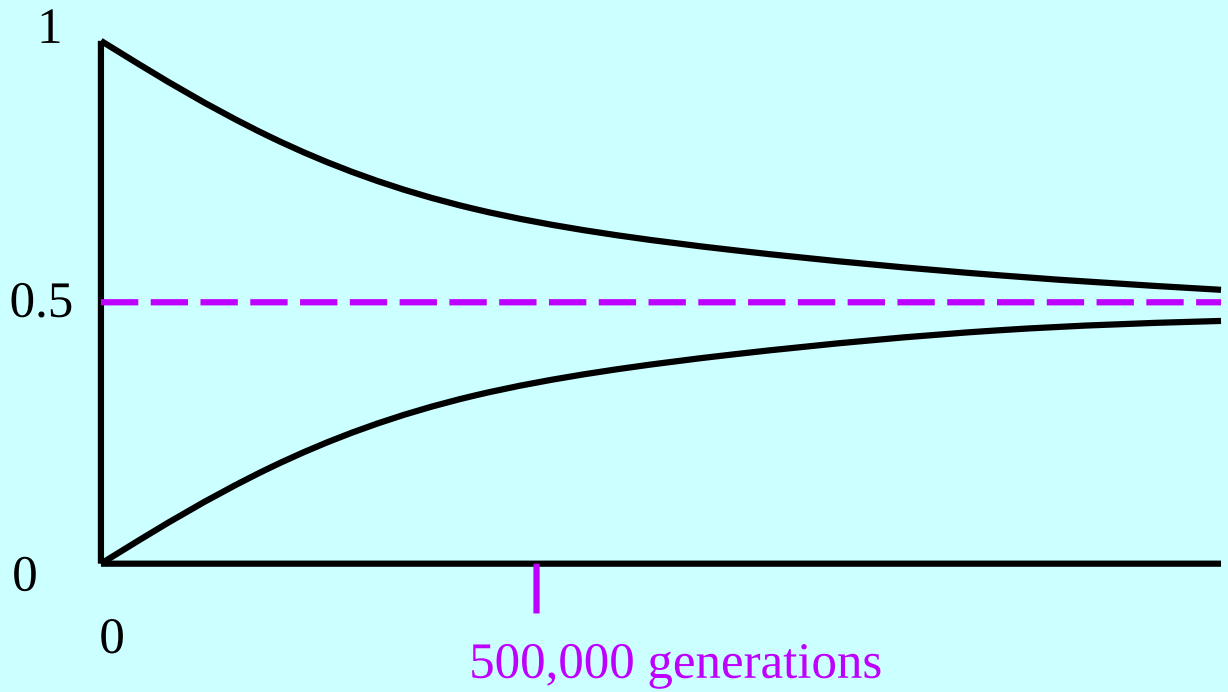
W	O	R	D
W	O	R	E
G	O	R	E
G	O	N	E
G	E	N	E

But the word BCGH cannot be made into an English word

There are also only a tiny fraction of all 456,976 four-letter words that are English words But they are clustered, so that it is possible to “evolve” from one to another through intermediates.

Mutation as an evolutionary force

If we have two alleles A and a, and mutation rate from A to a is 10^{-6} and mutation rate back is the same,



Mutation is critical in introducing new alleles but is very slow in changing their frequencies

Estimation of a human mutation rate

By an equilibrium calculation. Huntington's disease. Dominant. Does not express itself until after age 40. $1/100,000$ of people of European ancestry have the gene. Reduction in fitness maybe 2%.

- If allele frequency is q, then $2q(1-q)$ of everyone are heterozygotes. So $q \simeq 0.000005$.
- 0.02 of these die. So a fraction 0.02 of all copies of the mutant allele in the population are eliminated by natural selection each generation.
- So the fraction of all gene copies at that locus that are mutant alleles that are eliminated is $0.000005 \times 0.02 \simeq 10^{-7}$
- If we are at equilibrium between mutation and selection, this is also the fraction of all gene copies at that locus that have a new mutation.
- So the mutation rate is in that case 10^{-7}

Similar calculations can be done with recessive alleles, but we must remember that in their case each death (or reduction in fitness) kills two copies of the mutant.