

Selection and evolution A2

Read online at pastopic.com/cambridge/biology-9700/notes/selection-and-evolution

Last checked 30 September 2026

Read first: [Inheritance](#)

© 2026 Pastopic. Free for personal study. Please share the link, not the file.

What you need to know

By the end of this topic you should be able to:

1. **Explain, with examples, that phenotypic variation comes from genes, the environment, or both.**
2. **Explain what discontinuous and continuous variation are.**
3. **Explain the genetic basis of each.**
4. **Use the t -test to compare the means of two samples** (formula given).
5. **Explain why natural selection happens:** many offspring compete for resources, and in the "struggle for existence" the best adapted are most likely to survive, reproduce and pass on their alleles.
6. **Explain how the environment acts as a stabilising, directional or disruptive force of selection.**
7. **Explain how selection, the founder effect and genetic drift (including the bottleneck effect) change allele frequencies.**
8. **Outline how bacteria become resistant to antibiotics.**
9. **Use the Hardy–Weinberg principle** to calculate allele and genotype frequencies, and state when it applies (equations given).
10. **Describe the principles of selective breeding** (artificial selection).
11. **Outline three examples:** disease resistance in wheat and rice; inbreeding and hybridisation for vigorous, uniform maize; higher milk yield in dairy cattle.
12. **Outline the theory of evolution:** new species form from existing ones over time as gene pools change.
13. **Discuss how DNA sequence data show relationships between species.**
14. **Explain how speciation happens by genetic isolation:** geographical separation (allopatric) or ecological and behavioural separation (sympatric).

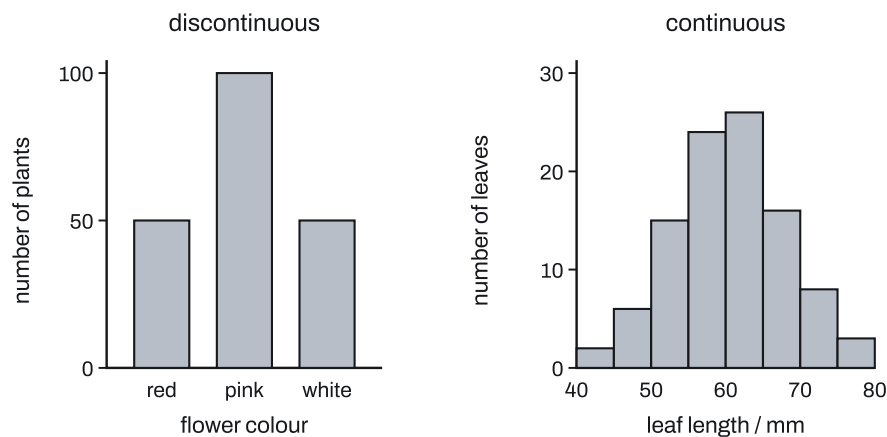
Understand it

Where variation comes from

Differences in **phenotype** have three possible causes.

- **Genetic** (V_G). New alleles come from **mutation**, a random change in DNA; sexual reproduction reshuffles alleles by **crossing over** and **independent assortment** in meiosis, **random mating** and **random fertilisation**. ABO blood groups depend only on genes.
- **Environmental** (V_E): food, light, temperature, water, minerals, disease. Clones (plants from cuttings of one parent) have identical alleles, so their differences are environmental.
- **Both** ($V_G + V_E$), for most features: genes set the possible range, the environment decides where in it an individual ends up. A person's mass depends on many genes and on diet.

Discontinuous and continuous variation



	Discontinuous	Continuous
Phenotypes	a few distinct classes, no in-betweens	a range between two extremes, no distinct classes
Genes	one gene, or a few	many genes (polygenes) at different loci
Effect of each allele	large: one allele can change the class (dominant, recessive, codominant)	small, and the effects add together (additive)
Environment	little effect	large effect
Graph	bar chart with gaps	histogram, often a bell-shaped (normal) curve
Examples	blood group, flower colour in many plants	height, mass, milk yield, leaf length

Why polygenes give a range. If three genes each have an allele that adds 2 cm to height and one that adds nothing, a plant can have 0 to 6 "adding" alleles: seven classes, 2 cm apart. The environment blurs the steps into a smooth range, and most individuals sit in the middle.

Comparing two samples: standard deviation and the t -test

The **standard deviation** (s) measures the spread of the data around the mean. Two samples nearly always have different means by chance; the **t -test** asks whether the difference is **significant** (too large to be chance). Use it for continuous data that are roughly normally distributed.

1. **Null hypothesis**: there is no significant difference between the two means.
2. Calculate t with the given formula (below).
3. **Degrees of freedom** = $n_1 + n_2 - 2$. Look up the critical value at $p = 0.05$.
4. t **greater** than the critical value: the difference is significant (less than 5% probability it is due to chance), so reject the null hypothesis. Smaller: not significant.

A significant difference does not prove its **cause**: ask whether other variables were controlled.

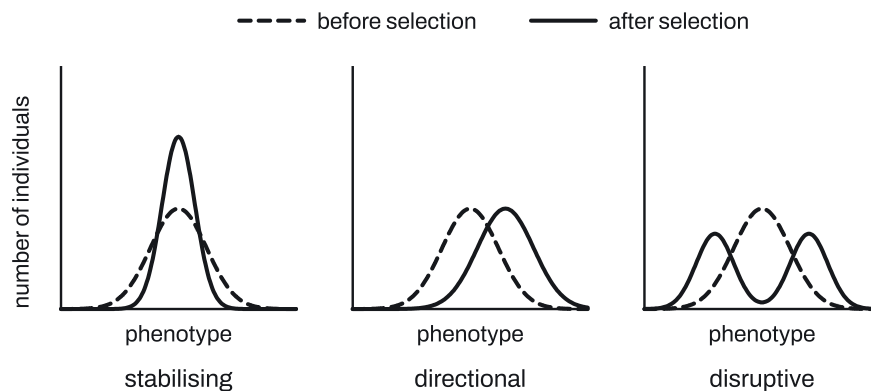
Natural selection

1. Organisms produce **more offspring** than the environment can support.
2. The offspring **vary**, partly because of their alleles.
3. They **compete** for limited resources such as food, space or mates, and face predators and disease: the **struggle for existence**. Most die before they breed, which is why population sizes stay fairly steady.
4. Anything in the environment that affects survival is a **selection pressure**. Individuals whose alleles give a phenotype suited to it have a **selective advantage**: they are more likely to **survive and reproduce**.
5. They pass on the advantageous **alleles**, so the **allele frequency** rises over many generations. The population becomes adapted.

Mutations are **random**: a selection pressure does not cause the mutation it favours, it only makes the allele more common once it exists.

- A new **dominant** advantageous allele spreads faster than a recessive one, because it shows in heterozygotes; a recessive allele is hidden until it is homozygous, which is rare while the allele is rare.
- A **short generation time** or **high mutation rate** speeds adaptation: more generations and more new alleles for selection to act on.

Three patterns of selection



- **Stabilising:** the environment stays the **same**. Both extremes are selected against, so the mean stays and the spread narrows (human birth mass). A species in an unchanging habitat can stay the same for millions of years. Selection favouring heterozygotes keeps both alleles in the population.
- **Directional:** the environment **changes** so one extreme is favoured, and the mean shifts that way (after a drought leaves only large seeds, birds with larger beaks survive; antibiotic resistance).
- **Disruptive:** both extremes are favoured and intermediates selected against (food in only two sizes, a habitat with two backgrounds). Two peaks form, which can lead to sympatric speciation.

Genetic drift, the founder effect and the bottleneck effect

Allele frequencies also change by **chance**, with no link to how well adapted an allele is.

- **Genetic drift:** random changes in allele frequency between generations, because it is chance which individuals mate or die and which gametes fuse. The frequency wanders up and down. In a **small population** the effect is large: an allele can be **lost** or **fixed** (frequency 1) in a few generations.
- **Founder effect:** a few individuals start a new population (say on an island). Their alleles are not representative of the original population, so some are missing and some rare ones become common.
- **Bottleneck effect:** a population crashes (disease, hunting, disaster) and the few survivors carry only some of the alleles; diversity stays low even if numbers recover.

All three **shrink the gene pool** and **raise homozygosity**; small populations also **inbreed**. **Migration** (gene flow) brings new alleles in. A difference that clearly helps survival is explained by selection; one with no obvious advantage, in a small population, may be drift.

Antibiotic resistance

1. A random **mutation** gives one bacterium resistance (say an enzyme that breaks down the antibiotic).
2. The **antibiotic is the selection pressure**: susceptible bacteria die, resistant ones survive and divide.
3. Offspring inherit the allele and its frequency rises fast: **directional selection**.

4. Resistance genes on **plasmids** pass between bacteria by **horizontal gene transfer: conjugation** (plasmid copied from cell to cell), **transformation** (DNA taken up from the surroundings) and **transduction** (carried by a virus). This spreads resistance much faster than mutation alone.

Stopping a course early leaves partly resistant bacteria alive to multiply. Sequencing bacterial DNA tracks resistance genes and guides the choice of antibiotic. A drug that blocks **translation** of the resistance gene's mRNA stops the resistance protein being made, so the antibiotic works again.

The Hardy–Weinberg principle

For a gene with two alleles, A (frequency p) and a (frequency q):

$$p + q = 1 \qquad p^2 + 2pq + q^2 = 1$$

p^2 , $2pq$ and q^2 are the frequencies of AA , Aa and aa . Only aa individuals show the recessive phenotype, so you can count them: **start from q^2** , then $q = \sqrt{q^2}$, $p = 1 - q$, and $2pq$. Every value lies between 0 and 1.

The principle says allele frequencies stay **constant** from generation to generation, but only if:

- the population is **large** (so no genetic drift);
- mating is **random**;
- there are **no mutations**;
- there is **no migration** in or out (no gene flow);
- there is **no selection** (every genotype survives and breeds equally well).

If the frequencies **do** change between generations, a condition is broken. For example, if people with a recessive disease die younger, fewer aa people are alive to count, so q^2 , and so $2pq$, come out too low.

Selective breeding (artificial selection)

Humans act as the selection pressure:

1. choose parents with the **desired characteristics** (features, not alleles);
2. cross them;
3. select the **best offspring** and breed from those;
4. **repeat for many generations**, until the offspring reliably show the feature.

Genetic diversity falls. **Inbreeding** (breeding close relatives) makes organisms homozygous and brings harmful recessive alleles together (**inbreeding depression**: less vigour and fertility); **outbreeding** unrelated varieties avoids this.

Disease resistance in wheat and rice. Cross a variety with a resistance allele (often low-yielding) with a high-yielding one. Keep offspring that resist the disease and yield well, and cross them back to the high-yielding parent for several generations (backcrossing), selecting for resistance each time, so the new variety has the yield plus the resistance allele. DNA sequencing checks the allele is present. Short stems (dwarf plants don't fall over) or flood tolerance are brought in the same way.

Vigorous, uniform maize.

1. **Inbreed** two lines for several generations, selecting for desired traits. Each becomes **homozygous**: uniform but weak.
2. **Hybridise** the two lines. Remove the **tassels** (male flowers) of the female parent so it cannot self-pollinate.
3. The **F1** plants are heterozygous at many loci and genetically identical: **vigorous** (hybrid vigour) and **uniform** (same height, ripen together, easy to harvest).
4. Their seed gives a **variable** F2 (not true-breeding), so farmers buy new F1 seed each year.

Milk yield in dairy cattle. Record each cow's yield. Breed the best cows with bulls whose mothers and daughters were high yielders: bulls make no milk, so a bull is judged by his daughters (**progeny testing**). Artificial insemination lets one bull father thousands of calves. Select the best daughters and repeat for many generations. It is slower than with crops: long generation time, one calf at a time, and a bull is proven only years later.

	Natural selection	Selective breeding
Selection pressure	the environment	humans
Mating	random	chosen by humans
Features favoured	those that help survival and reproduction	those humans want, which may harm the organism
Speed	usually many generations	faster
Genetic diversity	usually kept	reduced; inbreeding likely

Genetic engineering is faster still: it can add an allele in one generation rather than over many.

Evolution and species

The **theory of evolution**: over many generations gene pools change (by selection and drift), phenotypes change, and new species form from pre-existing ones. A **species** (biological species concept) is a group that can interbreed to produce **fertile offspring**; a new species forms when populations become **reproductively isolated**.

DNA sequence data and evolutionary relationships

1. Sequence the same DNA (a gene, or mitochondrial DNA) from each species.
2. Line the sequences up and **count the base differences**, using databases and software (bioinformatics).
3. Mutations build up over time, so **fewer differences means more closely related**: a more recent common ancestor.
4. If mutations build up at a roughly constant rate (a **molecular clock**), the number of differences estimates how long ago two species diverged, and the results can be drawn as a branching tree.

Mitochondrial DNA comes from the mother only and does not cross over, so it changes only by mutation. Compared with morphology (bones, teeth), DNA gives more information, is **quantitative** and precise, needs only a small sample, and is not fooled by unrelated species that look alike. Divergence dates stay **uncertain**: the mutation rate is an estimate, old DNA breaks down, and fossils are few.

Speciation

Allopatric speciation (geographical separation):

1. A **geographical barrier** (sea, river, canyon, ice sheet) splits one population into two.
2. There is **no gene flow** between them.
3. Different **random mutations** arise in each.
4. **Different selection pressures** (climate, food, predators) select different alleles; drift also acts in small populations.
5. The **allele frequencies and gene pools diverge**, and so do the phenotypes.
6. Eventually they cannot interbreed to produce fertile offspring even if they meet: **reproductively isolated**, two species.

Sympatric speciation happens with **no geographical separation**. Gene flow stops because of **ecological separation** (different habitats or foods in the same area, so the groups rarely meet) or **behavioural separation** (different courtship, mating calls or breeding seasons). Disruptive selection often drives it; then allele frequencies diverge until the groups are reproductively isolated.

Key facts and formulas

"Given" means it is printed on the paper when needed. "Learn it" means you must remember it.

Formula	Status
Hardy–Weinberg: $p + q = 1$ and $p^2 + 2pq + q^2 = 1$	Given
$t = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Given
Degrees of freedom for the t -test = $n_1 + n_2 - 2$	Learn it
$t > \text{critical value at } p = 0.05$: significant difference	Learn it
$q = \sqrt{\text{frequency of the recessive phenotype}}$, then $p = 1 - q$	Learn it
Number of heterozygotes = $2pq \times \text{population size}$	Learn it

How to do it

We have tagged 58 Paper 4 questions for this topic, 2021–2025 (49 different: 9 are repeated on Paper 41). Two use moths only as the setting for a genetics question, so the table counts the other 56; 37 of those mix types, so the counts add up to more than 56.

Question type	Questions
Explain speciation (allopatric or sympatric)	21
Continuous and discontinuous variation and their genetic basis	12
Selective breeding (maize, wheat and rice, dairy cattle)	12
Explain how natural selection changes a population	11
Genetic drift, the founder effect and the bottleneck effect	11
DNA sequence data and evolutionary relationships	11
Read evolution data from a graph or table	9
Genetic and environmental causes of variation	8
Stabilising, directional and disruptive selection	7
Hardy–Weinberg calculations and conditions	6
Antibiotic resistance	3
The <i>t</i> -test and standard deviation	2

1. Explain speciation (21 questions)

1. **Decide which kind.** A barrier on the map → **allopatric**. Groups in the **same area** differing in food, habitat or behaviour → **sympatric** (say "no geographical separation").
2. Write the steps in context: separation → no gene flow → different random mutations → different selection pressures (named) → different allele frequencies → cannot produce fertile offspring. Name the type.
3. Say **populations** until the last step, and **alleles**, not genes.

Variations: examples of behavioural separation; the biological species concept; why two different-looking forms are one species (they produce fertile offspring together); an 8-mark essay, which needs every step plus drift in small populations.

2. Continuous and discontinuous variation (12 questions)

1. **Name the type:** a range or histogram → continuous; separate bars → discontinuous.
2. **Genetic basis:** use the table in "Understand it" (many genes, small additive effects, plus environment; or one gene, large effects).
3. In tick tables, "environment may affect gene expression" applies to both.

3. Selective breeding (12 questions)

1. Give the principles: choose parents with the feature, cross, select the best offspring, repeat for many generations.
2. Add the steps for the named example in order (maize: inbreed → homozygous lines → hybridise → uniform, vigorous F1; tassels removed; new seed each year).

Variations: compare with natural selection or genetic engineering; why cattle are harder than crops; disadvantages (lost diversity, inbreeding depression); why backcross.

4. Explain how natural selection changes a population (11 questions)

1. **Variation:** a random mutation gave a new allele.
2. **Name the selection pressure** from the story (a disease, a predator, a drought).
3. Individuals with the allele have a **selective advantage**: they **survive and reproduce** more.
4. They **pass on the allele**, and its **frequency increases** over many generations.
5. Name the pattern if it fits, usually **directional selection**.

Link each step to the context. **Variations:** why populations stay stable; why a dominant allele spreads faster; why short generations speed adaptation.

5. Genetic drift, the founder effect and the bottleneck effect (11 questions)

1. Say it is **chance**, not selection.
2. Give the mechanism: which individuals mate or die, which gametes fuse (drift); a few individuals start a new population (founder); a crash leaves few survivors (bottleneck).
3. Give the result: alleles lost or fixed, rare alleles become common, homozygosity rises, smaller gene pool.
4. Say the effect is **larger in small populations**.

Variations: a rare phenotype common on an island (founder effect, drift, inbreeding; selection only if it helps); list every process that changes allele frequencies (mutation, selection, drift, founder, bottleneck, migration); selection or drift?; bringing in individuals raises diversity (gene flow).

6. DNA sequence data (11 questions)

1. Sequence the same gene; count base differences (bioinformatics).
2. Fewer differences → more closely related, more recent common ancestor; a steady mutation rate dates the split (molecular clock).
3. Apply to the figure: name the closest pair.

Variations: advantages over comparing bones and teeth; why a date of divergence is uncertain; outline the theory of evolution; a fill-in-the-gaps passage.

7. Read evolution data from a graph or table (9 questions)

1. **Describe**: the trend ("as X increases, Y decreases") with figures. For two alleles, the other allele's frequency is $1 -$ the plotted one.
2. **Explain** with selection or drift (more disease → survivors carry the resistance allele → that phenotype becomes common).

8. Genetic and environmental causes of variation (8 questions)

1. Clones → **environmental** only; distinct types set by genes (blood groups) → **genetic** only; named genes **plus** named conditions → **both**.
2. "Describe the causes": genetic, environmental and their interaction, **each with an example**.

9. Stabilising, directional and disruptive selection (7 questions)

1. **Name it**: unchanged environment → stabilising; one extreme favoured → directional; both extremes → disruptive.
2. **Explain**: the pressure, the phenotypes selected for and against, the allele-frequency change.
3. **Sketch** as in the diagram: narrower (same peak); whole curve moved (start, peak and end); two peaks.

10. Hardy–Weinberg (6 questions)

1. Find q^2 as a decimal (6.8% → 0.068).
2. If you are given the **dominant** phenotype (say 40% resistant), then $q^2 = 1 - 0.40 = 0.60$.
3. $q = \sqrt{q^2}$; $p = 1 - q$; $2pq$.
4. Multiply by the population size if a **number** is asked; give a percentage if a percentage is asked. Show every line.

Variations: name the principle; say which conditions a population meets and why (an isolated island: no migration; a poison: selection).

11. Antibiotic resistance (3 questions)

Give the four steps in "Understand it" in order; for "two ways" to become resistant, give mutation and horizontal gene transfer.

12. The t -test (2 questions)

1. Substitute into the formula; show the numerator and the denominator.
2. Give t to the accuracy asked (4 significant figures means 4.255, not 4.26).
3. Compare with the critical value: " t is greater than the critical value, so the difference between the means is significant and not due to chance." State which mean is higher.

Worked examples

The bold codes show where marks are earned, as in Cambridge mark schemes. The number is how many marks.

- **M1** method mark: for a correct method, even if the numbers slip.
 - **A1** accuracy mark: for a correct answer; it needs the M mark before it.
 - **B1** mark for a correct result on its own, with no method needed.
-

Example 1

In a population of 2400 moths, wing colour is controlled by one gene. Pale wings are recessive; 9% of the moths have pale wings.

(a) Use the Hardy–Weinberg equations to calculate the number of moths that are heterozygous. **[3]**

(b) Give two conditions needed for the Hardy–Weinberg principle to apply. **[2]**

(a) $q^2 = 0.09$, so $q = \sqrt{0.09} = 0.3$. **B1**

$p = 1 - 0.3 = 0.7$ and $2pq = 2 \times 0.7 \times 0.3 = 0.42$. **B1**

Number of heterozygotes = $0.42 \times 2400 = 1008$. **B1**

(b) Any two, **B1** each: a large population; random mating; no mutation; no migration into or out of the population; no selection for or against any genotype.

Example 2

A population of lizards once lived across a single large island. About 500 000 years ago, rising sea level split the island into two, with 20 km of sea between them. The eastern island is dry with sparse plants; the western island is wet and forested. Today the lizards on the two islands are two species.

Explain how the two species could have evolved. **[5]**

- The sea is a **geographical barrier** splitting the population in two. **B1**
- There was **no gene flow** between them. **B1**
- Different **random mutations** arose in each population. **B1**
- There were **different selection pressures** (dry and open against wet and forested, so different food, predators and camouflage), so different alleles were selected; the **allele frequencies** of the two gene pools became different. **B1**
- Eventually the populations became **reproductively isolated**: they cannot interbreed to produce fertile offspring. This is **allopatric speciation**. **B1**

Example 3

A student measured the leaf length of plants of one species growing in full sun and in shade.

	n	mean / cm	standard deviation / cm
sun	20	6.4	1.1
shade	18	7.3	1.4

(a) Calculate the value of t . Give your answer to 2 decimal places. [2]

(b) The critical value at $p = 0.05$ is 2.03. State what the student can conclude. [2]

(a)

$$t = \frac{7.3 - 6.4}{\sqrt{\frac{1.1^2}{20} + \frac{1.4^2}{18}}} = \frac{0.9}{\sqrt{0.0605 + 0.1089}} = \frac{0.9}{0.4116} = 2.19$$

M1 for the correct substitution, **A1** for 2.19.

(b) $2.19 > 2.03$, so there is a **significant difference** between the means and it is not due to chance. **B1**

Leaves in the shade are longer. **B1**

Example 4

A population of grasshoppers lives on a hillside of green grass. A fire leaves the hillside as patches of green grass and black burnt ground, with few areas in between. Birds eat grasshoppers they can see. Grasshopper body colour ranges from light green to black.

Name the type of selection now acting, and explain how it will change the population over many generations. [4]

Disruptive selection. **B1** Predation by birds is the selection pressure. **B1** Green grasshoppers on grass and black grasshoppers on burnt ground are camouflaged, so they are more likely to survive and reproduce; intermediate colours stand out on both and are eaten. **B1** The frequencies of the alleles for green and for black rise, and the population forms two peaks of colour. **B1**

Common mistakes

- **Confusing genetic drift with natural selection.** Drift is chance and does not adapt a population; "selection pressures caused genetic drift" is wrong. They are two separate mechanisms.
- **Saying the selection pressure caused the mutation.** Mutations are random. The pressure only decides which existing alleles become more common.
- **Calling the separated groups "species" from the start.** They are populations until they can no longer interbreed.

- **Writing "genes" when you mean "alleles"**. Selection changes allele frequencies; continuous variation is many genes, each with alleles of small effect.
- **Stopping too early**. Add the selection pressure and the allele-frequency change; end speciation answers with "cannot produce fertile offspring".
- **Assuming allopatric speciation** when the figure shows both groups in one area.
- **Hardy–Weinberg slips**: using the recessive phenotype frequency as q instead of q^2 ; forgetting to multiply $2pq$ by the population size; getting a value above 1.
- **t -test slips**: forgetting the square root in the denominator, and rounding in the middle.
- **Selective breeding**: breeders select features, not alleles, then select and breed the best offspring; bulls are chosen by their daughters' yield.
- **Antibiotics**: the **bacteria** become resistant, not the patient (never "immune").

How to get the marks

- **"Explain"** is marked "any four from" a list: each correct point in context earns a mark, so write several short linked points using the key terms (selection pressure, selective advantage, allele frequency, gene flow, reproductive isolation).
- **Use the context**: name the barrier, pressure and phenotype from the question.
- **"Describe" a graph**: the trend in words plus figures from the graph, with units.
- **"Suggest"**: apply your knowledge to a new situation. **"Outline"**: the main steps in order. **"State"/"Name"**: a word or phrase ("directional").
- **Calculations**: show each step (working earns marks) and give the accuracy asked.
- **Sketches**: a smooth curve on the given axes with the change obvious.
- **Compare**: both sides of each point ("natural selection is ..., whereas selective breeding is ...").

Quick check

1. The seed mass of a bean variety ranges from 0.4 g to 1.1 g with no distinct groups. Name this type of variation and explain its genetic basis.
2. In a population of plants, 64% have the dominant phenotype (tall). Calculate the percentage of plants that are heterozygous.
3. A population of 5000 birds on the mainland has a rare recessive allele for white feathers. Six birds reach an empty island and start a new population, in which white birds become common. Explain why.
4. Suggest why it takes longer to raise the milk yield of dairy cattle by selective breeding than to raise the grain yield of wheat.
5. The same 800-base gene was sequenced in three species. Species P and Q differ at 12 bases; P and R differ at 44 bases. Calculate each difference as a percentage, and state which species is most closely related to P.

Answers

1. **Continuous variation.** Seed mass is controlled by many genes (polygenes) at different loci; each allele has a small effect and the effects add together. The environment also affects it, giving a smooth range.
2. Recessive (short) = $1 - 0.64 = 0.36 = q^2$, so $q = 0.6$ and $p = 0.4$. $2pq = 2 \times 0.4 \times 0.6 = 0.48$, so **48%** are heterozygous.
3. **Founder effect:** the six founders are not a representative sample of the mainland alleles; by chance they carried the white allele at a much higher frequency. In such a small population, **genetic drift** can raise its frequency further, and inbreeding brings two recessive alleles together, so more birds are homozygous and white.
4. Cattle take about two years to reach breeding age and have one calf at a time, so each generation is slow and small. Only cows produce milk, so a bull's value is known only from his daughters' milk yield (progeny testing), years later. Wheat has a generation every year, makes many seeds, and every plant shows its own yield.
5. P and Q: $\frac{12}{800} \times 100 = 1.5\%$. P and R: $\frac{44}{800} \times 100 = 5.5\%$. **Q** is most closely related to P: fewer base differences mean less time for mutations to build up since their common ancestor.

Past questions to try

- [9700/43/O/N/25 Q4](#): natural selection by disease, with a graph to interpret.
- [9700/42/M/J/25 Q2](#): genetic drift and the founder effect.
- [9700/42/O/N/24 Q2](#): the processes that change allele frequencies, and maize breeding.
- [9700/42/O/N/25 Q4](#): antibiotic resistance.