REVISION GUIDE

When a gene *changes*

Mutation and gene editing

One base out of place can change a protein, or change nothing at all. Edit a gene yourself and watch the protein, follow sickle cell from one base to a bent red blood cell, find out where mutations come from, and at higher level switch genes off, edit them with CRISPR and ask why some sequences never change.

9 chapters · a dry lab

The interactive guide, with every animation:

biologybybradford.com/guides/ib-biology-d1-3-mutation-and-gene-editing

Contents

01	Three ways to change a gene	3
02	One base swapped	5
03	Sickle cell, one base at a time	8
04	Bases added or lost	12
05	Where mutations come from	14
06	Passed on, or not	17
07	Switch a gene off HL	19
08	Cut, then repair HL	21
09	Sequences that stay the same HL	25
	Key terms	27
	Dry lab: Conserved or not?	28
	Check yourself	29
	Answers	30

By the end you can

- [ ] Distinguish between substitutions, insertions and deletions as changes to the base sequence of a gene D1.3.1
- [ ] Explain why a base substitution may or may not change an amino acid, and what a SNP is D1.3.2
- [ ] Explain why insertions and deletions are likely to stop a polypeptide working, through a frameshift or a large change D1.3.3
- [ ] Outline the causes of gene mutation: mutagens, with chemical and radiation examples, and errors in replication or repair D1.3.4
- [ ] Explain in what sense mutation is random, and in what sense it is not D1.3.5
- [ ] Compare the consequences of mutation in germ cells and in somatic cells D1.3.6
- [ ] Explain why mutation is the original source of all genetic variation, and why it matters for evolution D1.3.7
- [ ] Outline how knocking out a gene is used to find out what it does D1.3.8
- [ ] Describe how CRISPR sequences and Cas9 are used to edit genes, with one successful use D1.3.9
- [ ] Evaluate two hypotheses for why some sequences are conserved across species D1.3.10

Tick each one when you can do it without the book.

01 Three ways to *change* a gene

D1.3.1 GENE MUTATIONS AS STRUCTURAL CHANGES

A gene is a sequence of bases, read in threes. A **gene mutation** is a change to that sequence. It is structural: the DNA molecule itself is different, base by base, and every copy made from it carries the change.

There are three kinds. In a **substitution** one base is swapped for another. In an **insertion** one or more bases are added. In a **deletion** one or more bases are lost.

The three look alike as edits, but they do very different things to the protein. A substitution changes at most one codon. An insertion or a deletion moves every base after it along, so every codon after it can change. Step through the figure, then try the same edits on a gene in chapter 2.

KEY POINTS

- A gene mutation is a change to the base sequence of a gene.
- Substitution: one base replaced. Insertion: bases added. Deletion: bases lost.
- A substitution keeps the gene's length; insertions and deletions change it.

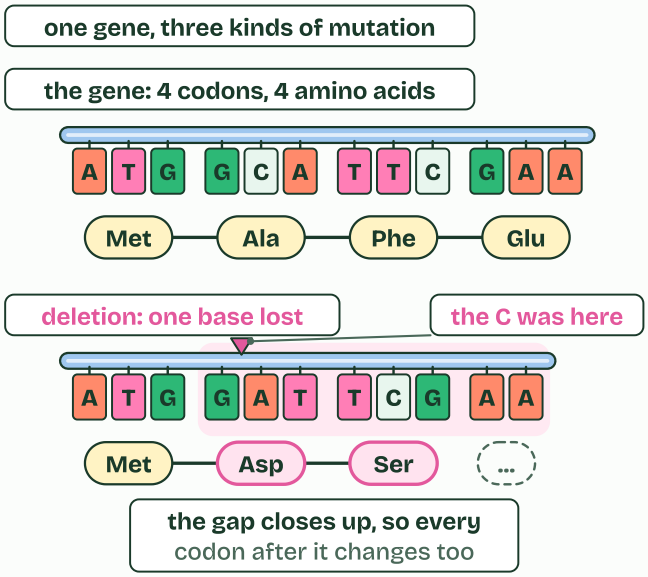
KEY TERMS

gene mutation
A change to the base sequence of a gene. It is a change to the structure of the DNA, at the level of the molecule.

substitution
A gene mutation in which one base is replaced by a different base. The length of the gene does not change.

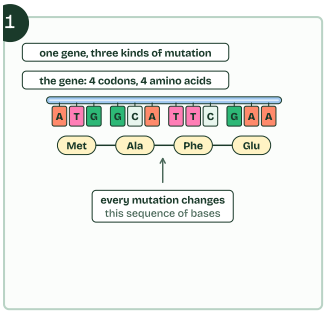
insertion
A gene mutation in which one or more extra bases are added to the sequence.

deletion
A gene mutation in which one or more bases are lost from the sequence.

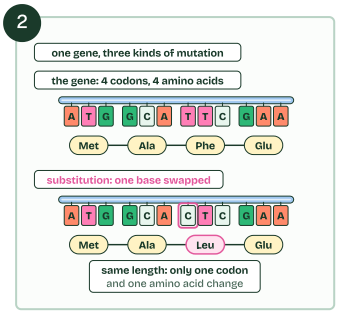


STEP BY STEP

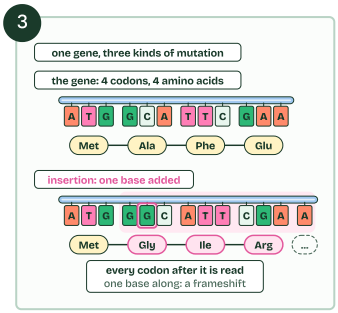
4 STEPS



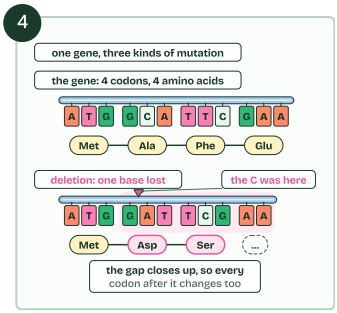
A short gene: twelve bases, read as **four codons**.



A **substitution**: the T at base 7 becomes a C. Only the third codon changes.



An **insertion**: one G added after base 4. Every base after it moves along one.



A **deletion**: base 5 lost. The bases close up, and every codon after it is new.

TRY IT FIRST

Read this sentence in threes: THE OLD MAN SAW THE DOG. Now swap one letter for another. Then remove one letter and read on in threes. Which edit wrecks more of the sentence? Chapter 4 comes back to this.

WATCH OUT

A mutation is not a whole chromosome going missing, and it is not "a gene turning bad". It is any change to the bases, and many changes do nothing that anyone could notice.

PREDICT

A gene of 900 bases becomes 901 bases long. Which kind of mutation could this be?

(A) A substitution, because one of the 900 bases has been swapped for a different base

(B) An insertion, because one extra base has been added somewhere in the sequence

(C) A deletion, because one base has been lost and the gap closed up on both sides

Decide before you look. Answer on page 30.

02 One base *swapped*

D1.3.2 CONSEQUENCES OF BASE SUBSTITUTIONS

A substitution changes one codon. What happens next depends on what that codon now codes for, and there are three possible results.

The genetic code is **degenerate**: most amino acids have more than one codon, and the codons for one amino acid usually differ only at the third base. So a substitution can be **silent**: the codon changes, the amino acid does not, and the polypeptide is exactly as it was. If the new codon codes for a different amino acid, the mutation is **missense**: one amino acid in the chain is different. If the new codon is a stop codon, it is **nonsense**: translation ends early and the polypeptide is cut short.

A base substitution that is found in some members of a population is a **single-nucleotide polymorphism (SNP)**. Any two people differ at millions of SNPs. Because the code is degenerate, a SNP may or may not change an amino acid, and most of them lie outside genes altogether.

KEY POINTS

- A substitution changes one codon, and so at most one amino acid, unless it makes a stop codon.
- Silent: same amino acid, because the code is degenerate. Missense: a different amino acid. Nonsense: a stop codon.
- SNPs are the result of base substitutions; they may or may not change an amino acid.

KEY TERMS

degenerate
Having more than one codon for most amino acids. 61 codons code for 20 amino acids, so most amino acids have two, four or six codons.

single-nucleotide polymorphism (SNP)
A position in the genome where a single base differs between members of a species. SNPs are the result of base substitution mutations.

Each codon: DNA coding strand, then mRNA, then amino acid

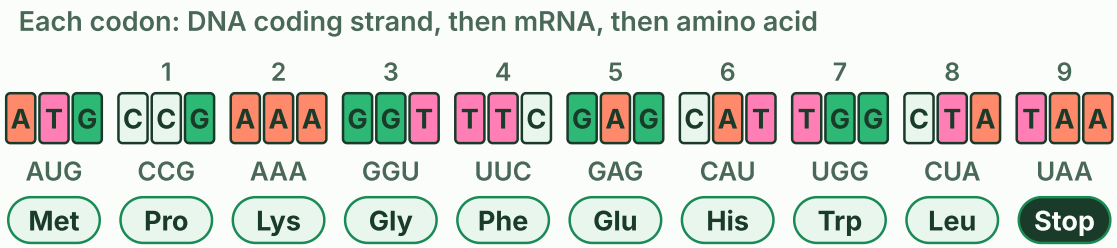
ATG	CCG	AAA	GGT	TCG	AGC	ATT	GGC	TAT	AA
Met	Pro	Lys	Gly	Ser	Ser	Ile	Gly	Tyr	

A deletion: one base lost, the frame shifts

Tap a base, then change it, add a base after it or delete it. The tool names the result. Try the five tasks.

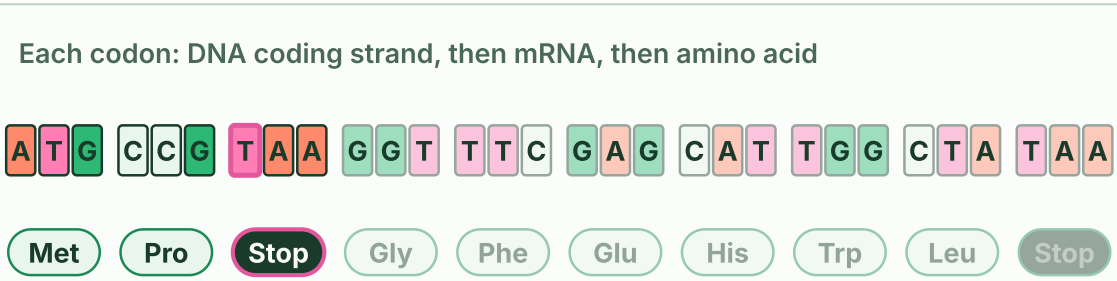
STEP BY STEP
3 STEPS

1



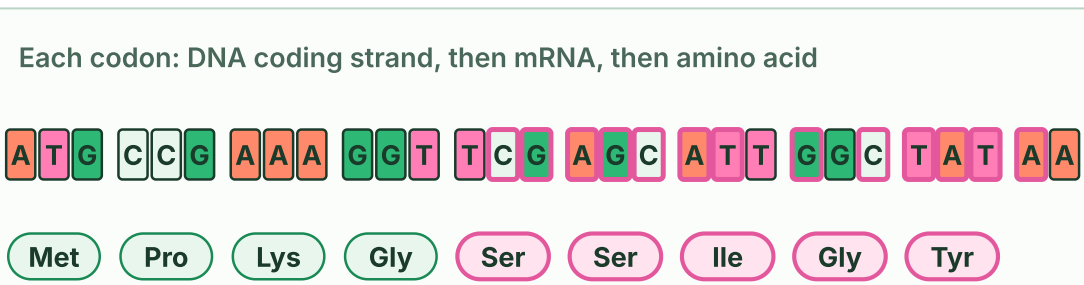
A made-up gene: its coding strand in codons, the mRNA under it and the amino acid each codon codes for.

2



One base **substituted**: codon 2, AAA, becomes TAA, a stop codon. Translation ends there: a **nonsense** mutation.

3



One base **deleted** from codon 4. Every codon after it is read in a new frame, so every amino acid from there changes: a **frameshift**.

REMEMBER IT

What a swapped base can do

S Same	M Mistake	S Stop
Silent new codon, same amino acid	**Missense** one amino acid different	**Nonsense** a stop codon: chain cut short

WATCH OUT

"Missense" does not always mean harmful. Swapping one amino acid for a similar one, on the outside of a protein, often changes nothing you could measure. A change in the active site of an enzyme can stop it working.

PREDICT

The codon GGU codes for glycine, and so do GGC, GGA and GGG. The U is replaced by an A. What is the result?

(A) A missense mutation, because the base sequence of this codon is now different

(B) A silent mutation, because GGA still codes for glycine

(C) A nonsense mutation, because the change makes a stop codon at this point

Decide before you look. Answer on page 30.

03 Sickle cell, *one base* at a time

D1.3.2 A SUBSTITUTION WITH A KNOWN EFFECT: SICKLE CELL

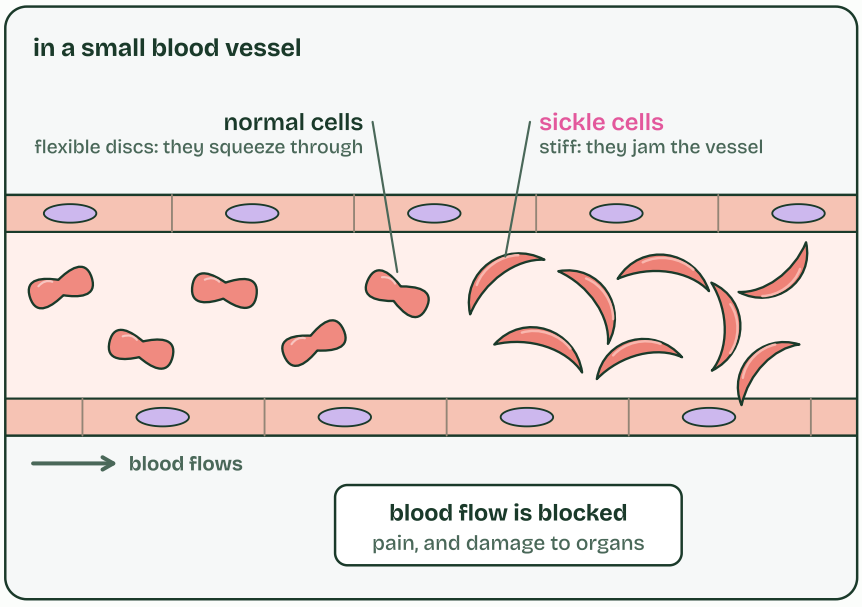
Haemoglobin, the oxygen-carrying protein in red blood cells, is built from two kinds of polypeptide. One of them, beta globin, is coded for by a gene on chromosome 11. The usual allele, Hb^A, has GAG as its sixth codon (counting from after the start codon), and GAG codes for **glutamic acid**.

The sickle cell allele, Hb^S, differs by one base: the A in that codon is replaced by a T. GAG becomes GTG, and GTG codes for **valine**. It is a missense substitution, and one amino acid out of 146 is the whole difference.

That one amino acid matters because of its R group. Glutamic acid's is charged and hydrophilic; valine's is non-polar and hydrophobic, so it sits on the outside of the molecule as a sticky patch. When oxygen is low, haemoglobin S molecules stick to one another and form long, stiff fibres. The red blood cells bend into a sickle shape, carry less oxygen, and block small blood vessels. People with two copies of Hb^S have sickle cell disease.

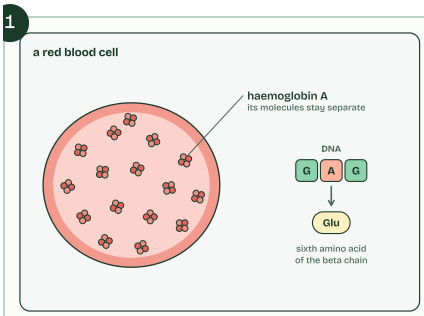
KEY POINTS

- The sickle cell allele has a base substitution: GAG becomes GTG in the sixth codon of beta globin.
- Glutamic acid (charged) is replaced by valine (non-polar).
- Haemoglobin S forms fibres at low oxygen, so red blood cells sickle.



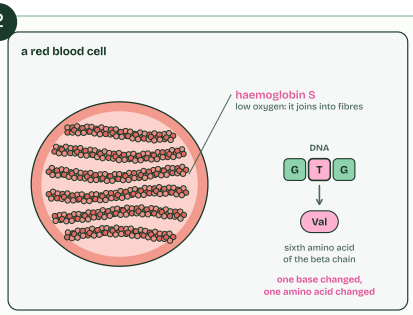
STEP BY STEP
3 STEPS

1



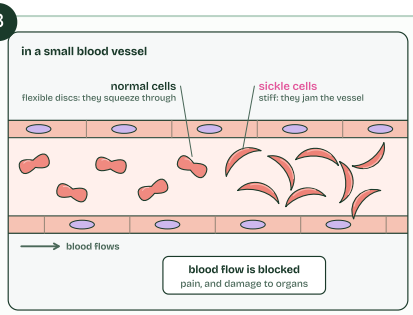
In a red blood cell, molecules of **haemoglobin A** stay separate.

2

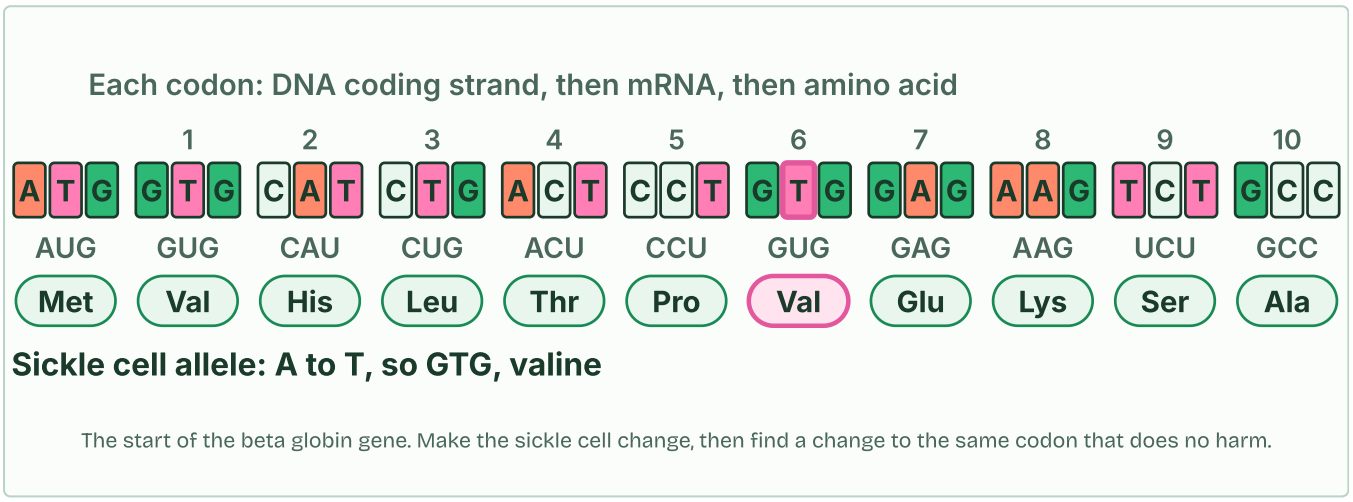


Molecules of **haemoglobin S**, at low oxygen, stick together into long fibres.

3

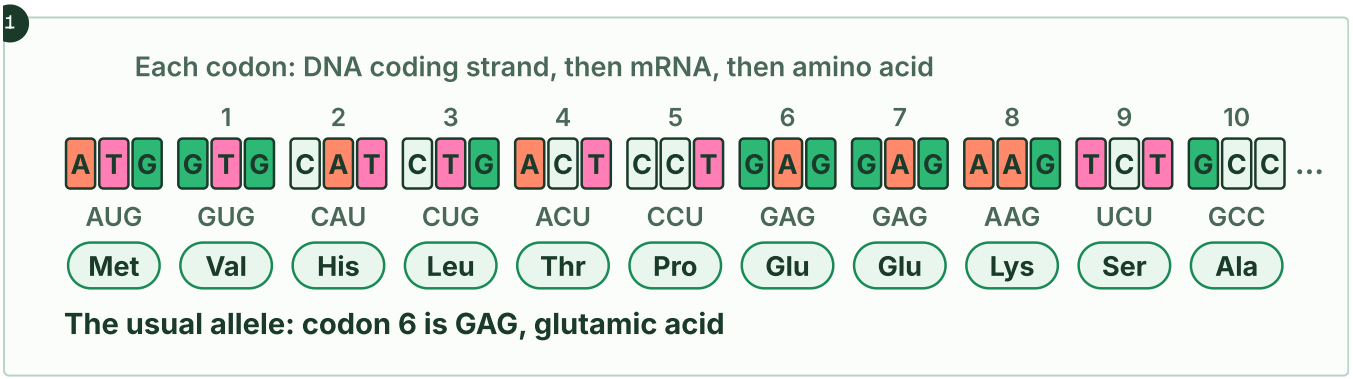


The fibres bend the cell into a **sickle**. Sickle cells are stiff and block small vessels.



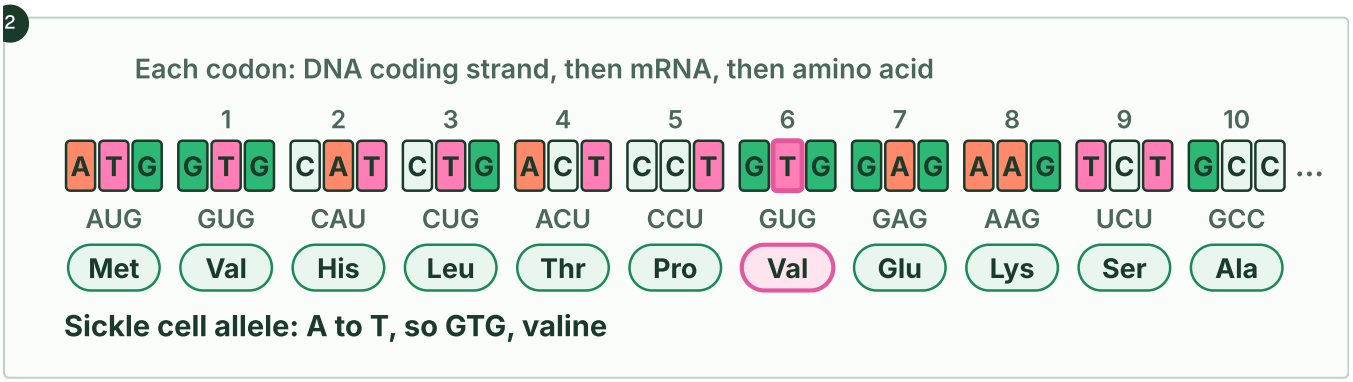
STEP BY STEP
2 STEPS

1

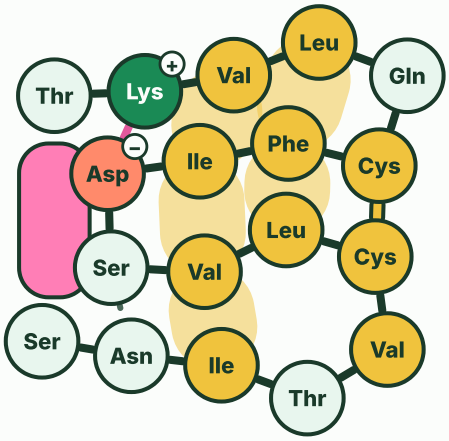


The start of the beta globin gene. Codon 6 is **GAG**, which codes for glutamic acid.

2



One base substituted, **A to T**: GAG becomes GTG, and glutamic acid becomes **valine**. That one amino acid is the whole difference between the two alleles.

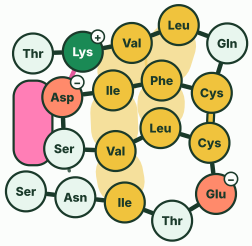


Make the sickle cell substitution in codon 6, A to T, and watch the amino acid change on the surface of a folded protein. Then try G to A, the silent change. It is a toy model of a made-up chain, not a real structure of beta globin.

STEP BY STEP

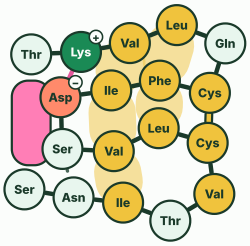
2 STEPS

1



A folded chain with glutamic acid (Glu, charged) on its surface, codon GAG. Nothing sticks to it.

2



Codon GTG, valine: the surface now carries a non-polar patch, **sticky** to the same patch on another copy. The fold inside is unchanged.

LINK

This is protein structure (B1.2) seen from the gene: the sequence of amino acids decides how a polypeptide folds and what it can do. One substitution changes the sequence in one place, and the protein's behaviour follows.

WATCH OUT

The DNA changes in one base, not the whole gene, and the protein changes in one amino acid, not all of it. Write the change in both: the base (A to T) and the amino acid (glutamic acid to valine).

PREDICT

A different substitution in the same codon turns GAG into GAA. What would happen to the haemoglobin?

(A) It would form fibres, because every change in this one codon gives the sickle cell form

(B) It would not be made, because a change in this codon always makes a stop codon instead

(C) Nothing would change, because GAA also codes for glutamic acid, so the chain is the same

Decide before you look. Answer on page 30.

04 Bases *added* or lost

D1.3.3 CONSEQUENCES OF INSERTIONS AND DELETIONS

A ribosome reads mRNA in threes from the start codon, with nothing to mark where one codon ends and the next begins. That grouping into threes is the **reading frame**.

Insert or delete one base and the frame shifts. Every codon after the change is read in a new grouping, so every amino acid after it can change, and a stop codon usually turns up early, because three of the 64 codons are stops. This is a **frameshift**, and the polypeptide it makes almost never works.

Add or remove three bases, or any multiple of three, and the frame is kept. Whole codons are gained or lost and the rest of the chain is as before. That can still matter: one amino acid missing from a key place can stop a protein folding properly. Large insertions or deletions, of many bases, are also very likely to stop a polypeptide working, whatever the frame.

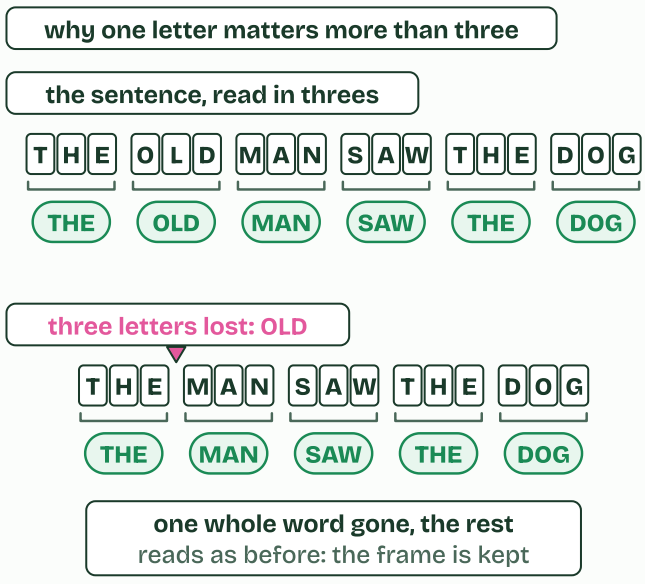
KEY POINTS

- Inserting or deleting one or two bases shifts the reading frame: a frameshift.
- After a frameshift every codon changes and an early stop codon is likely, so the polypeptide is very unlikely to function.
- A multiple of three keeps the frame, but whole amino acids are gained or lost; large insertions and deletions usually stop the polypeptide working.

KEY TERMS

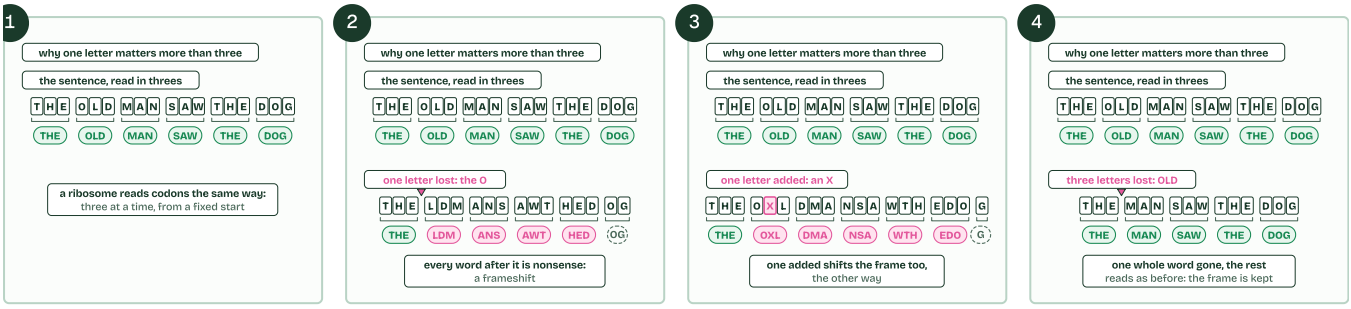
reading frame
The way a base sequence is divided into codons, set by the start codon. A shift of one or two bases changes every codon after it.

frameshift
A mutation that adds or removes a number of bases that is not a multiple of three, so every codon after it is read in a new grouping.



STEP BY STEP

4 STEPS



A sentence of three-letter words, read in threes like **codons**.

One letter lost. Read on in threes and every word after it is nonsense: a **frameshift**.

One letter added: the frame shifts the other way, and again every word after it changes.

Three letters lost: one whole word goes and the rest reads as before. The **frame is kept**.

WATCH OUT

The damage depends on where the change is, too. A frameshift near the end of a gene leaves most of the chain as it was; one near the start leaves almost nothing.

PREDICT

Three edits to the same gene: one base substituted, one base deleted, three bases in a row deleted. Which is most likely to stop the protein working?

(A) Three bases deleted, because it removes the most bases from the gene of the three

(B) One base deleted, because it shifts the frame, so every codon after it changes

(C) One base substituted, because it changes which amino acid sits at that position

Decide before you look. Answer on page 30.

05 Where mutations *come from*

D1.3.4 D1.3.5 CAUSES OF MUTATION, AND RANDOMNESS

Some mutations happen with no outside cause. DNA polymerase copies billions of bases each time a cell divides and, very rarely, puts in the wrong one. Proofreading and repair enzymes find and mend almost all of these errors, but not quite all, and repair can itself go wrong. So **errors in DNA replication or repair** are one cause.

The other is a **mutagen**. Chemical mutagens include benzo[a]pyrene and nitrosamines in tobacco smoke, aflatoxin made by a mould on stored nuts and grain, and mustard gas. Mutagenic radiation includes ultraviolet light, which can join two thymines next to each other, and the ionising radiation of X-rays, gamma rays and the alpha particles from radon gas, which can break DNA strands.

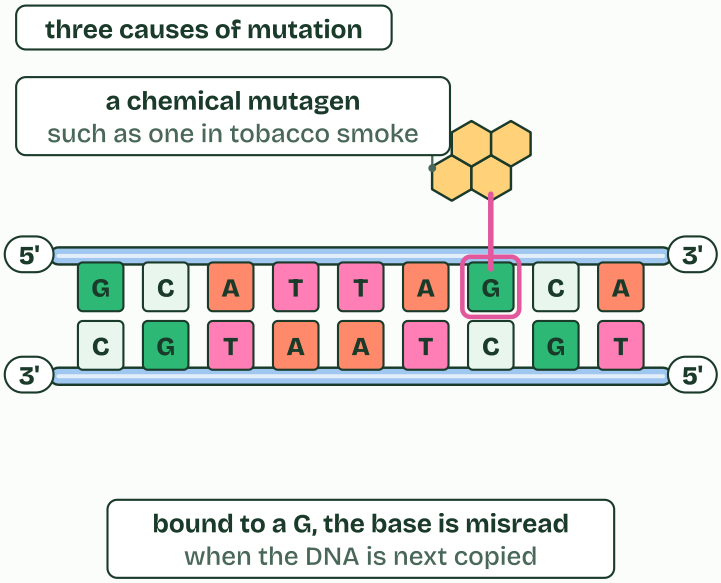
Mutation is **random** in two senses, and not in a third. It can happen anywhere in the base sequence of a genome. It happens regardless of whether the change would help: there is no known natural mechanism by which a cell makes a deliberate change to a particular base in order to change a trait. But not every base is equally likely to change: some sequences and some bases, such as a C next to a G, mutate more often than others.

KEY POINTS

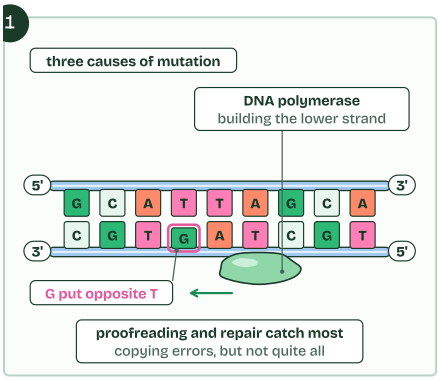
- Mutations are caused by errors in DNA replication or repair, and by mutagens.
- Mutagens: chemicals (such as those in tobacco smoke) and radiation (ultraviolet, X-rays, gamma rays, alpha particles).
- A mutation can occur anywhere, but some bases mutate more often than others; no natural mechanism makes a deliberate change to a chosen base.

KEY TERMS

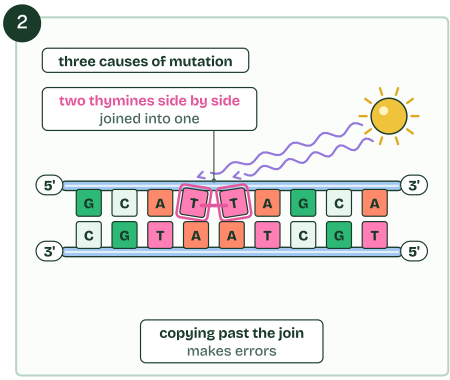
mutagen
Anything that increases the rate of mutation: some chemicals and some forms of radiation.



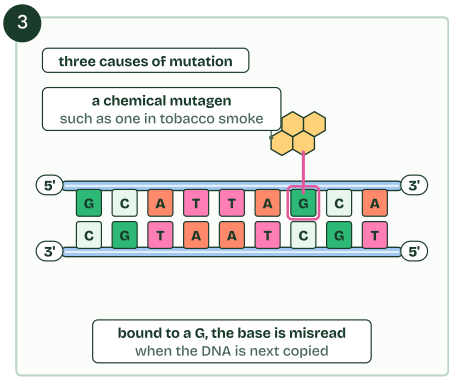
STEP BY STEP 3 STEPS



An **error in replication**: a G is put opposite a T. Repair enzymes catch most such errors.



Ultraviolet light joins two thymines next to each other. Copying past the join makes mistakes.



A **chemical mutagen** binds to a base, which is then misread when the DNA is copied.

REMEMBER IT

Mutagens

S Smoke	S Sun	S Scans
Chemicals	**Ultraviolet**	**Ionising radiation**
benzo[a]pyrene, nitrosamines, aflatoxin	joins two thymines	X-rays, gamma rays, radon

WATCH OUT

Bacteria exposed to an antibiotic do not mutate in order to resist it. The resistant mutants were already there, by chance, before the drug arrived; the drug kills the rest.

PREDICT

A student says: "A population of insects in a hotter climate will mutate so that it can cope with heat." What is wrong with this?

(A) Nothing is wrong: cells can sense heat and change the right base in the right gene to cope with it

(B) Insects do not have DNA that mutates, so the population could never change in response to heat at all

(C) Mutations happen at random with respect to need; heat does not cause the helpful ones

Decide before you look. Answer on page 30.

06 Passed on, or *not*

D1.3.6 D1.3.7 GERM CELLS, SOMATIC CELLS AND VARIATION

Every cell in your body descends from one zygote. Most are **somatic cells**, body cells. A few lines of cells, the **germ cells**, make the gametes.

A mutation in a somatic cell is copied into every cell descended from it, and nowhere else. Usually it does nothing. But if it hits genes that control cell division, that cell's descendants can divide without control and form a tumour: **cancer** begins with mutations in somatic cells, often several in one line of cells. A somatic mutation is not inherited: it ends with the person.

A mutation in a germ cell can end up in a gamete. If that gamete is used in fertilisation, the mutation is in the zygote, so it is in **every cell** of the child, germ cells included, and it can be passed on again. This is how a new allele enters a family.

Every allele of every gene began as a mutation, so mutation is the original source of all genetic variation. For one organism, most mutations are neutral or harmful, and few are beneficial. For a species, over a long time, they are essential: without new alleles there would be no variation for natural selection to act on, and no evolution (D4.1).

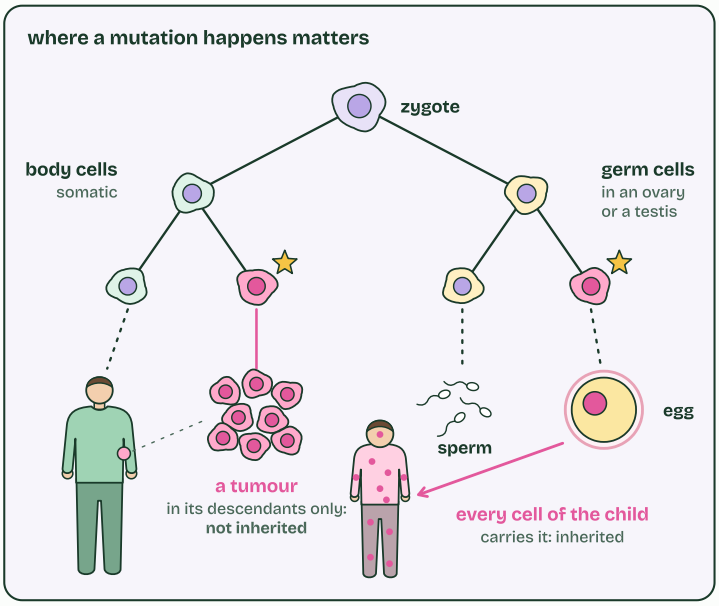
KEY POINTS

- A mutation in a germ cell can be inherited: it is in every cell of the offspring.
- A mutation in a somatic cell is not inherited; it can lead to cancer.
- Mutation is the original source of all genetic variation; most mutations are harmful or neutral, but they are essential for evolution.

KEY TERMS

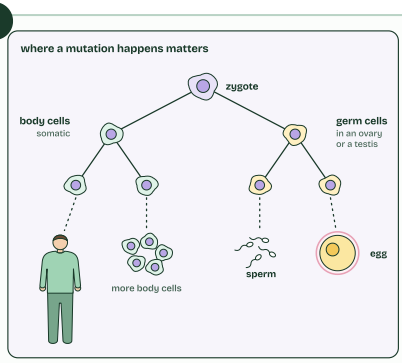
somatic cells
Any cell of the body that is not a germ cell. Mutations in somatic cells are not passed to offspring.

germ cells
Cells that give rise to gametes, eggs or sperm. A mutation in a germ cell can be passed to the next generation.



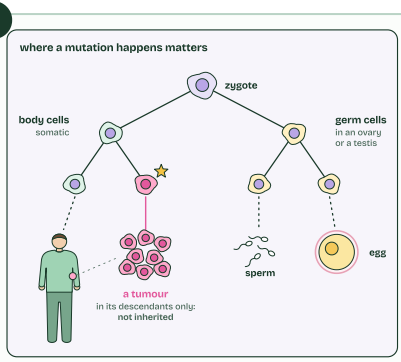
STEP BY STEP
3 STEPS

1



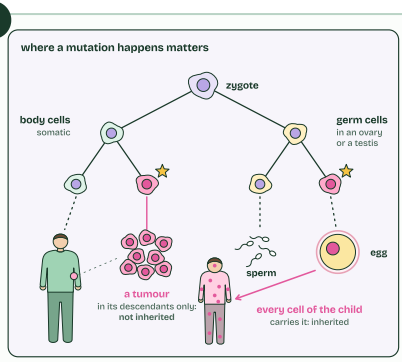
All the cells of a body descend from one **zygote**: somatic cells on one side, germ cells on the other.

2



A mutation in a **somatic cell** is in its descendants only. It can lead to a tumour, and it is not inherited.

3



A mutation in a **germ cell** can reach an egg or a sperm, and so every cell of a child.

NATURE OF SCIENCE

Commercial genetic tests read a person's SNPs from a saliva sample and report risks of future disease. Without expert interpretation the results can mislead: a raised risk is not a diagnosis, a test may look only at a few variants of a gene, and a clear result can reassure someone who still has other risks.

WATCH OUT

"Most mutations are harmful" is about one organism at a time. For a species, mutation is the only way new alleles arise, so without it natural selection would run out of variation to act on.

PREDICT

A man has a mutation in a lung cell after years of smoking. Could his children inherit it?

(A) No, because a lung cell is a somatic cell, and only germ cells make the gametes

(B) Yes, because every mutation in a person's body is copied into all the cells they have

(C) Yes, but only if the mutation goes on to cause cancer in that lung before the children are born

Decide before you look. Answer on page 30.

07 Switch a gene *off*

D1.3.8 HL GENE KNOCKOUT

Sequencing a genome tells you where its genes are, not what they do. One way to find out is to break a gene on purpose. In a **gene knockout**, a gene is changed so that it can no longer make a working product. The knocked-out organism is then compared with a normal one. Whatever is different, in its development, its body, its behaviour or its chemistry, points to what the gene does.

Knockouts are made in **model organisms**: species that are easy to keep and breed and that share many genes with humans, such as mice, zebrafish, fruit flies, yeast and the plant *Arabidopsis*. For some of them, scientists have built **libraries of knockouts**: collections with one strain for each of thousands of genes, so that a researcher can order a strain with the gene they study already switched off. Yeast has one for nearly all of its genes, and there are large collections for mice and for *Arabidopsis*.

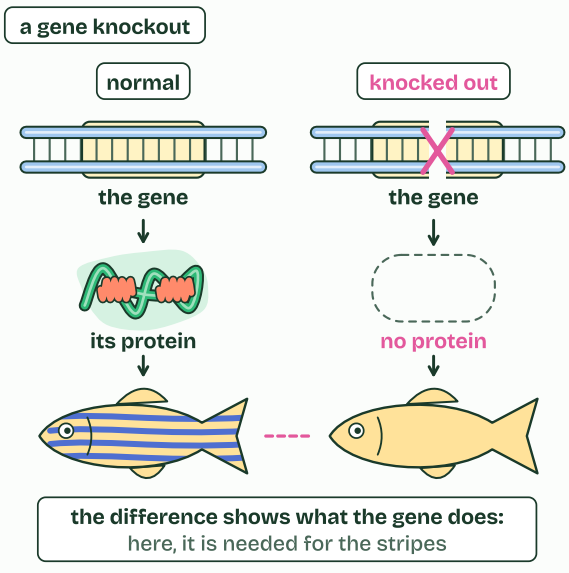
The methods have changed over the years, and you do not need to know them. CRISPR, in the next chapter, is now the usual way: a cut repaired with a few bases lost gives a frameshift, and the gene is knocked out.

KEY POINTS

- A knockout makes a gene inoperative, so that its function can be worked out from what changes.
- It is done in model organisms such as mice, zebrafish, yeast and *Arabidopsis*.
- Libraries of knockout organisms are available for some model species.

KEY TERMS

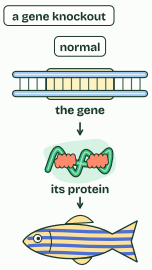
gene knockout
Making a gene inoperative, so that it no longer makes a working product, in order to find out what the gene does by comparing the organism with a normal one.



STEP BY STEP

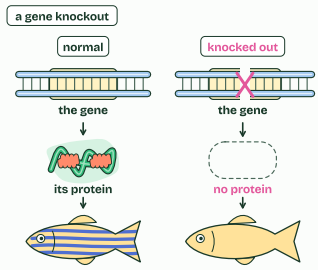
3 STEPS

1



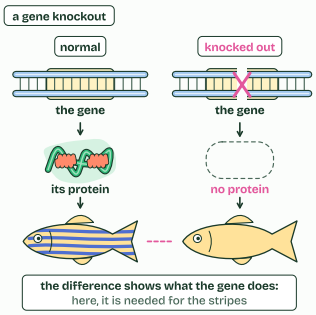
A **normal** gene makes its protein, and the fish develops its stripes.

2



The same gene **knocked out**: no working protein, and no stripes.

3



Compare the two. The difference shows **what the gene does**. The fish and the gene are drawn to show the idea.

WATCH OUT

No visible change does not mean the gene does nothing. Another gene may do the same job, or the effect may show only in certain conditions, such as cold, disease or old age.

PREDICT

A gene is knocked out in mice and the mice cannot smell. What is the best conclusion?

(A) The gene is the only gene that is needed for a mouse's sense of smell to work at all

(B) The gene's product is needed for a normal sense of smell in mice

(C) The gene codes for a protein found only in the nose, and is switched off everywhere else

Decide before you look. Answer on page 30.

08 Cut, then *repair*

D1.3.9 HL GENE EDITING WITH CRISPR AND CAS9

Cas9 is an enzyme that cuts DNA. It carries a **guide RNA**, and the guide RNA decides where it cuts. Scientists make a guide RNA with about 20 bases that match the gene they want to change. Cas9 moves along the DNA, opens the double helix where it finds a short sequence called the PAM, and checks whether the guide pairs. Where every base pairs, it cuts both strands. The system is named after the CRISPR sequences in bacteria, where it was found.

The cell then repairs the break, and this is where the change is made. Usually it simply joins the ends, often losing or adding a few bases: a frameshift that **knocks the gene out**. If scientists also supply a piece of DNA with the sequence they want, the cell can copy it across the break: an **edit**, such as one chosen base changed.

One successful use treats sickle cell disease. A patient's own blood stem cells are collected and edited with CRISPR-Cas9, outside the body, so that the cells make fetal haemoglobin again, which does not sickle. The cells are returned to the patient. Approved in the UK and the USA in 2023, the treatment has freed most patients in its trials from the painful crises of the disease.

For editing
Can cure diseases caused by one gene, such as sickle cell disease.
Faster, cheaper and more precise than earlier methods.
Knockouts reveal what genes do, in research.

Questions to settle first
Off-target cuts at sites that nearly match the guide.
Editing germ cells or embryos changes every later generation, who cannot consent.
Cost: who gets a treatment that costs over a million dollars?

KEY POINTS

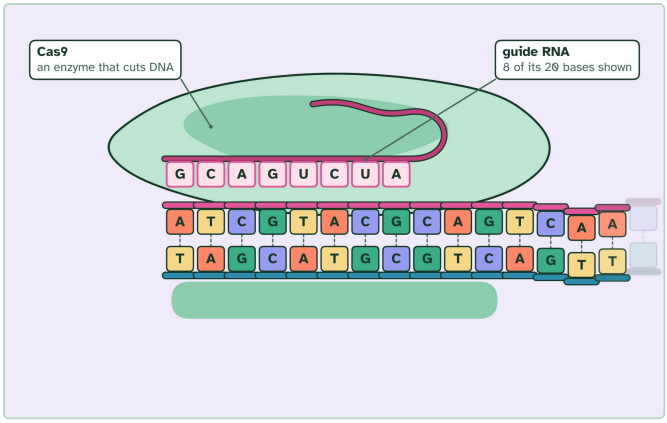
- A guide RNA matching the target leads Cas9 to it; Cas9 cuts both strands.
- Repair by joining the ends knocks a gene out; repair from a template makes a chosen edit.
- Example: edited blood stem cells treat sickle cell disease. Ethical issues and differing regulation need settling.

KEY TERMS

Cas9
An enzyme that cuts both strands of DNA at a site chosen by the guide RNA it carries.

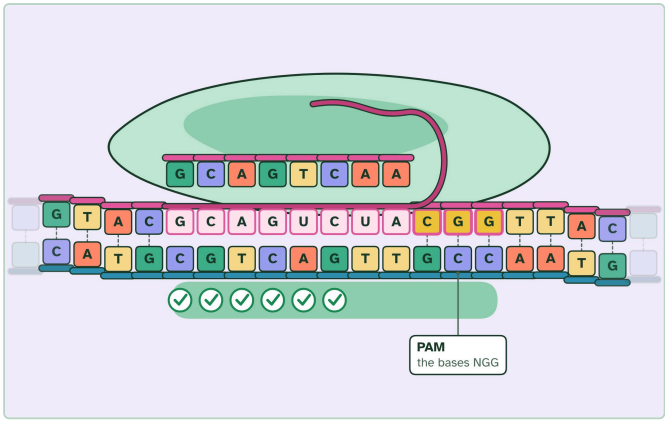
guide RNA
A short RNA made in the lab, with about 20 bases that match the stretch of DNA to be cut. It guides Cas9 to that site.

1 THE PARTS



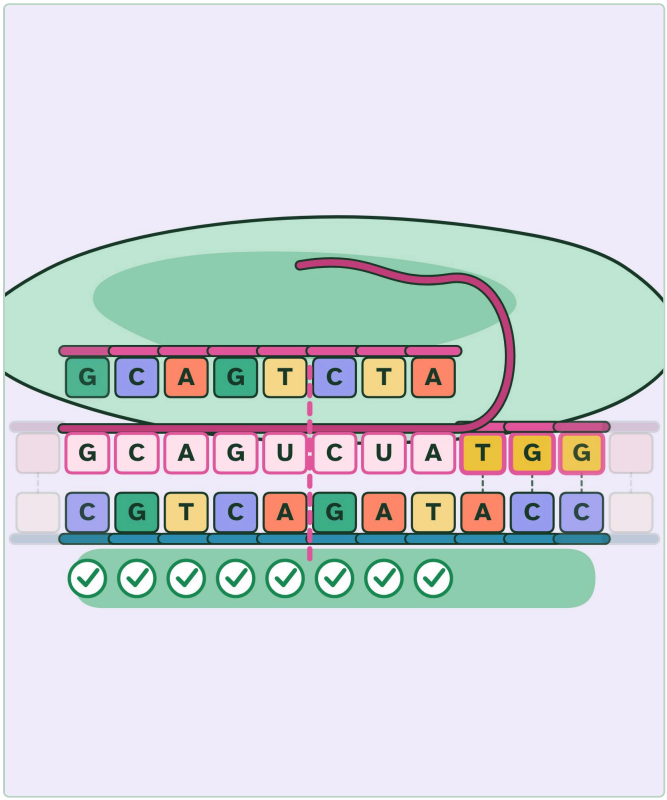
Cas9 is an enzyme that cuts DNA. It carries a **guide RNA**, made in the lab to match 20 bases of the gene to change. Eight of them are shown here.

2 A NEAR MISS



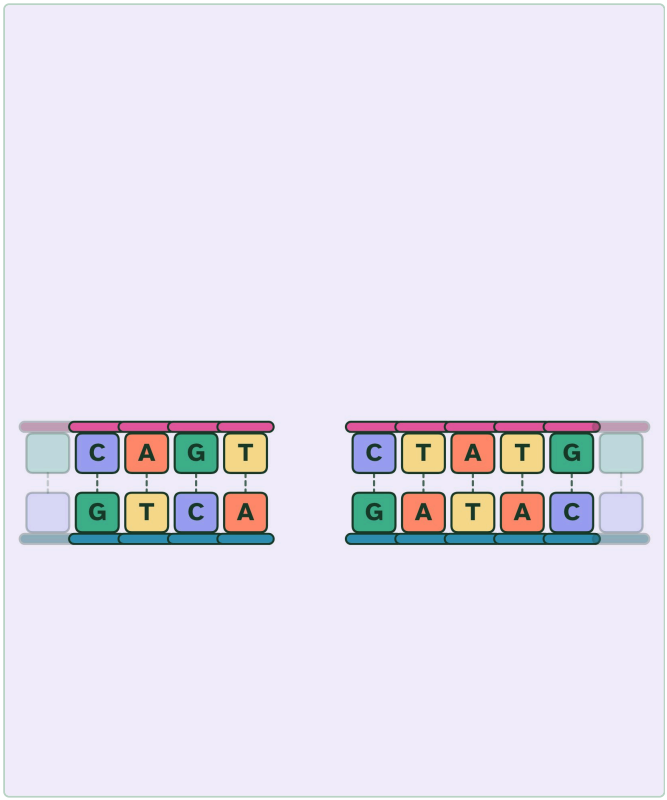
A near miss: one base does not pair, so Cas9 lets go. A site this close is a risk, because Cas9 sometimes cuts it: an **off-target** cut.

3 THE CUT



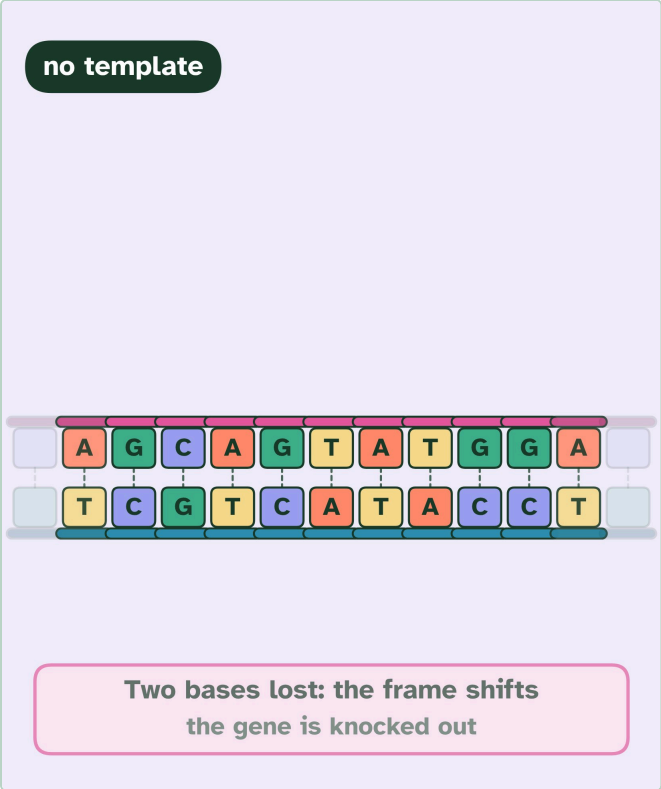
Every base of the guide pairs with the target, so Cas9 cuts **both strands** of the DNA, three bases from the PAM.

4 THE CUT



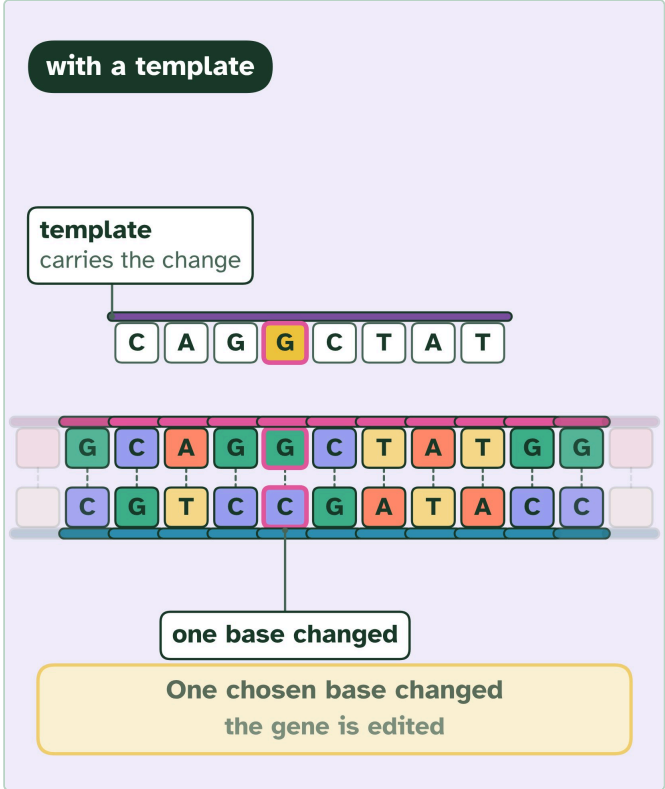
The DNA is broken, and Cas9 leaves. How the cell repairs the break decides the result.

5 KNOCKOUT



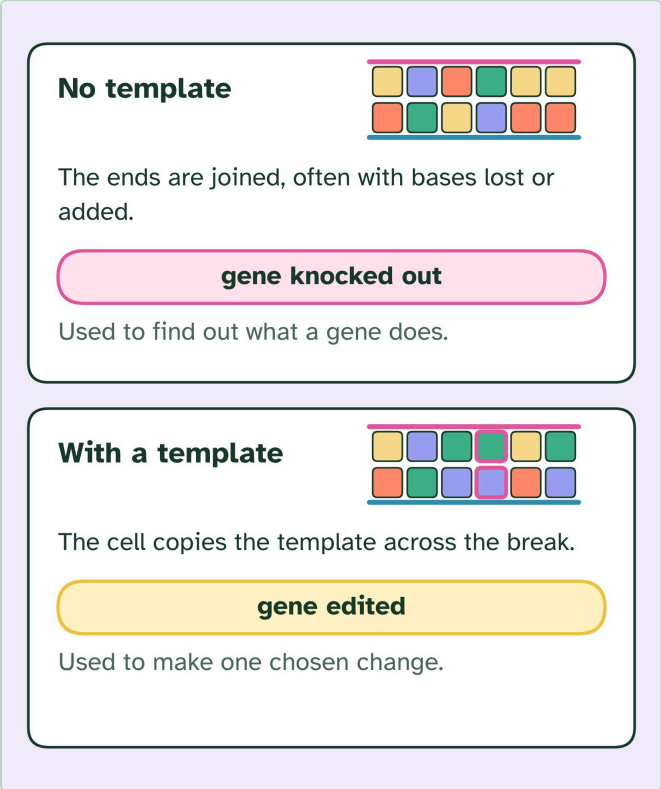
No template: the cell joins the ends, often losing or adding a few bases. Here two are lost, so the frame shifts and the gene is **knocked out**.

6 EDIT



The gene now carries the **chosen change**, one base, and nothing else.

7 TWO RESULTS



One cut, two results: switch a gene off, or edit it. Edited blood stem cells are now used to treat sickle cell disease.

Watch it move, and change it as it runs: biologybybradford.com/guides/ib-biology-d1-3-mutation-and-gene-editing#crispr

NATURE OF SCIENCE

Some uses of CRISPR raise ethical issues that must be settled before they are used. Countries regulate gene editing differently, and work banned in one may be allowed in another, so there is an international effort to harmonise the rules. In 2018 a scientist in China edited human embryos, and twins were born; he was widely condemned, and was jailed.

PREDICT

Scientists want to change one base in a gene, not knock the gene out. What must they add besides Cas9 and a guide RNA?

(A) A second guide RNA, so that Cas9 cuts out a whole section of the gene at both of its ends

(B) A template: DNA with the base they want, for the cell to copy as it repairs the cut

(C) Nothing, because Cas9 replaces the wrong base with the right one as it cuts the DNA strands

Decide before you look. Answer on page 30.

WATCH OUT

Cas9 does not "know" the gene. It cuts wherever the guide pairs next to a PAM, which is why a site that differs from the target by one or two bases can be cut by mistake.

09

D1.3.10 HL CONSERVED AND HIGHLY CONSERVED SEQUENCES

Sequences that *stay the same*

Line up the same gene from different species and most of it differs. Some stretches do not. A **conserved sequence** is identical or similar across a species or a group of species. A **highly conserved sequence** has stayed the same over long periods of evolution: the histone H4 protein of a pea and of a cow differ in only two of about a hundred amino acids.

Why would a sequence not change, when mutations happen everywhere? There are two hypotheses.

Function. The gene's product needs that sequence to work. A mutation there makes an organism less likely to survive and reproduce, so natural selection removes it, and the sequence stays as it was. Changes that do not alter the protein, such as most changes at a third codon base, are not removed.

Slower mutation. Some sequences may simply mutate less often, so fewer changes arise there in the first place.

The two predict different patterns, and that is how to test them. If function is the reason, silent changes still build up: third bases vary while the amino acids stay the same. If slower mutation is the reason, every kind of change is rare, silent ones included. A second test is a knockout or an edit: change a conserved stretch and see whether the protein still works. Try the first test in the Dry lab.

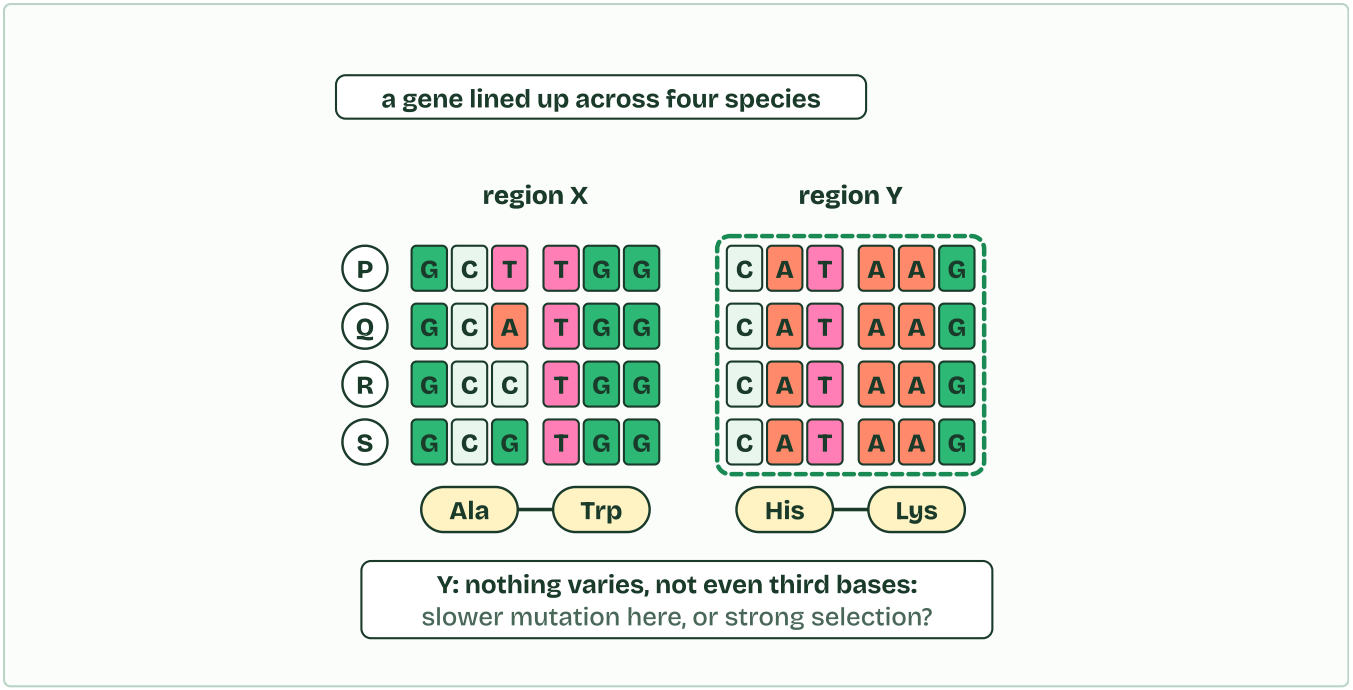
KEY POINTS

- Conserved: the same or similar across species. Highly conserved: the same over long periods of evolution.
- Hypothesis 1: the gene product's function needs the sequence, so selection removes changes.
- Hypothesis 2: the sequence mutates more slowly. Silent changes tell the two apart.

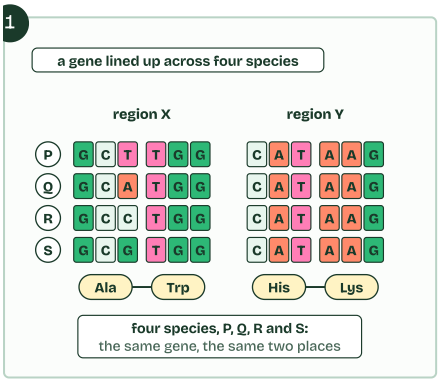
KEY TERMS

conserved sequence
A base or amino acid sequence that is identical or similar across a species or a group of species.

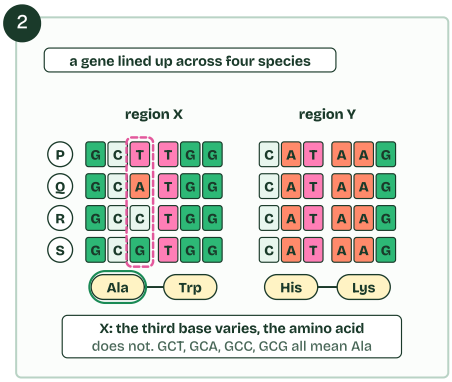
highly conserved sequence
A sequence that has stayed identical or similar over long periods of evolution, so it is shared by distantly related species.



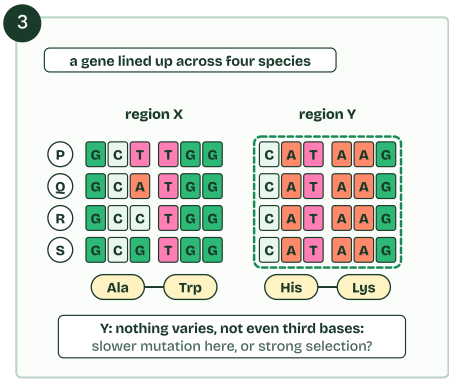
STEP BY STEP 3 STEPS



Two regions of one gene in four species, with the amino acids each codes for.



Region X: a **third base varies** but the amino acids do not. That fits the **function** hypothesis.



Region Y: **nothing varies**, not even third bases. That fits **slower mutation**, or very strong selection.

WATCH OUT

A conserved sequence has not been protected from mutation. Under the function hypothesis, mutations happen there as often as anywhere; the organisms carrying them are removed by selection.

PREDICT

In a conserved region of a gene, the amino acids are the same in ten species but the third bases of the codons differ. Which hypothesis does this support?

(A) Slower mutation, because the amino acids have not changed in any of the ten species here

(B) Function, because silent changes build up while changes to the amino acids do not

(C) Neither, because third bases never change in any gene, so they tell us nothing at all

Decide before you look. Answer on page 31.

Key terms

Every term this guide defines, A to Z, with the chapter it is in. Cover the right-hand side and test yourself.

Cas9 08
An enzyme that cuts both strands of DNA at a site chosen by the guide RNA it carries.

conserved sequence 09
A base or amino acid sequence that is identical or similar across a species or a group of species.

degenerate 02
Having more than one codon for most amino acids. 61 codons code for 20 amino acids, so most amino acids have two, four or six codons.

deletion 01
A gene mutation in which one or more bases are lost from the sequence.

frameshift 04
A mutation that adds or removes a number of bases that is not a multiple of three, so every codon after it is read in a new grouping.

gene knockout 07
Making a gene inoperative, so that it no longer makes a working product, in order to find out what the gene does by comparing the organism with a normal one.

gene mutation 01
A change to the base sequence of a gene. It is a change to the structure of the DNA, at the level of the molecule.

germ cells 06
Cells that give rise to gametes, eggs or sperm. A mutation in a germ cell can be passed to the next generation.

guide RNA 08
A short RNA made in the lab, with about 20 bases that match the stretch of DNA to be cut. It guides Cas9 to that site.

highly conserved sequence 09
A sequence that has stayed identical or similar over long periods of evolution, so it is shared by distantly related species.

insertion 01
A gene mutation in which one or more extra bases are added to the sequence.

mutagen 05
Anything that increases the rate of mutation: some chemicals and some forms of radiation.

reading frame 04
The way a base sequence is divided into codons, set by the start codon. A shift of one or two bases changes every codon after it.

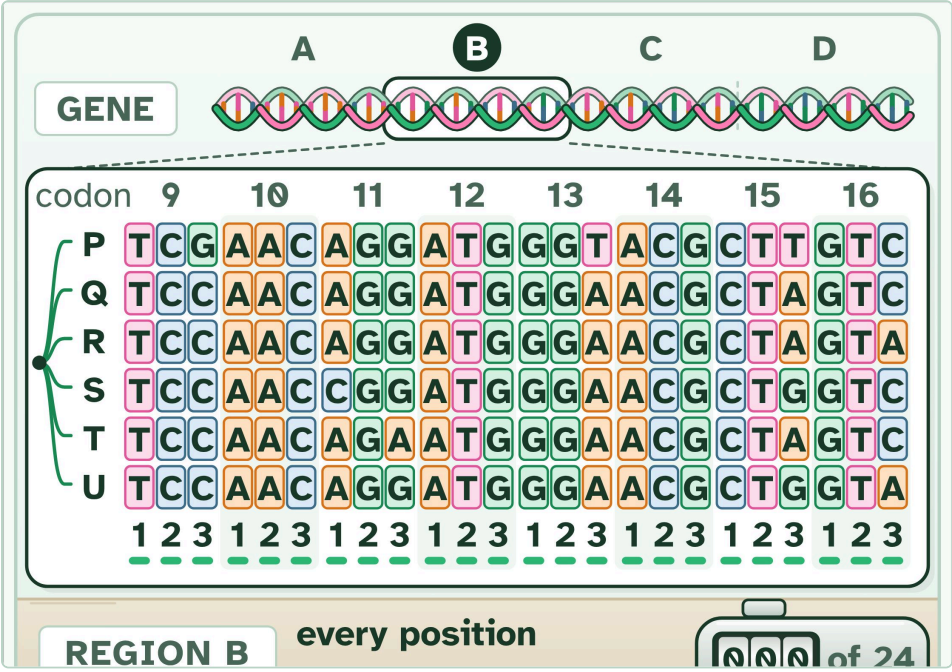
single-nucleotide polymorphism (SNP) 02
A position in the genome where a single base differs between members of a species. SNPs are the result of base substitution mutations.

somatic cells 06
Any cell of the body that is not a germ cell. Mutations in somatic cells are not passed to offspring.

substitution 01
A gene mutation in which one base is replaced by a different base. The length of the gene does not change.

Dry lab: *Conserved or not?*

One gene, lined up across six species. Find the stretches that never change, then test two hypotheses for why: does the protein need them, or do those bases simply mutate less? This bench is for higher level.



Run it online: change the variables, run the trials, and the table and graph fill in as you go. `biologybybradford.com/guides/ib-biology-d1-3-mutation-and-gene-editing#lab`

Your question

Why are some parts of a gene the same in every species? If the protein needs them, changes there are removed by selection, but silent changes should still build up. If those bases mutate less, every kind of change should be rare there. Which does the data support?

What you can change

The region of the gene you look at, the species you include, and the kind of position you count: every position, first and second codon positions (which mostly change the amino acid) or third positions (which mostly do not). The gene and species are invented to show the method.

What to record

For each region: the number of positions compared, the number that differ between the species, and the percentage, counted separately for the two kinds of position. Graph the two percentages for each region and decide which hypothesis each region supports.

Find out

- **Which regions are conserved?** Count every position in each of the four regions with all six species. Which regions differ least?
- **Function or slower mutation?** For each region, count first and second positions, then third positions. A region where only third positions vary fits the function hypothesis; one where neither varies fits slower mutation.
- **Look at the amino acids.** Show each region as amino acids. In which regions are the amino acids the same in every species, and are the bases the same there too?
- **Repeat with other species.** Count each region again with three species at a time (P, Q, R, then S, T, U). Do the regions keep their order? How much do the percentages scatter?

Record

- For every count: the region, the positions counted, the species included, the number of positions compared and the number that differ, and the percentage.
- At least three counts for each region and kind of position, each with a different set of species. The graph shows each count as a point and the mean as a bar.
- For each region, write which hypothesis the data support, and why. Say what other evidence would test it further.
- Download the CSV to keep your table. Standard level students can do the tasks in the gene editing tool in chapter 2 instead.

Your question Why are some parts of a gene the same in every species? If the protein needs them, changes there are removed by selection, but silent changes should still build up. If those bases mutate less, every kind of change should be rare there. Which does the data support?**What you can change**The region of the gene you look at, the species you include, and the kind of position you count: every position, first and second codon positions (which mostly change the amino acid) or third positions (which mostly do not). The gene and species are invented to show the method.**What to record**For each region: the number of positions compared, the number that differ between the species, and the percentage, counted separately for the two kinds of position. Graph the two percentages for each region and decide which hypothesis each region supports.

Back to the genetic code

Use the genetic code and natural selection to explain what your counts suggest about why regions A, B and C are, or are not, conserved.

QUESTIONS ON THE LAB

QUESTION 1

In region B the first and second positions hardly ever differ, but third positions often do. Which hypothesis do the data support?

(A) Slower mutation, because the first and second positions in region B almost never change

(B) Function, because mutations happen there but those that change amino acids are removed

(C) Neither, because a region in which some positions vary cannot be a conserved region

Answer on page 31.

QUESTION 2

Region C differs little at every kind of position, even at the third positions where region B varies freely. Why can the data not settle, on their own, why it is conserved?

(A) Because the silent third positions would vary if the protein needed it, and they do not

(B) Because slower mutation and very strong selection, even on silent changes, would look the same

(C) Because six species is too few to show any differences between the regions of a gene

Answer on page 31.

QUESTION 3

Why does a count with three species usually find fewer differing positions than the same count with six?

(A) A position counts as differing if any species differs, and with more species there are more chances for one to differ

(B) With fewer species the gene is shorter, so there are fewer positions in the region to compare

(C) Three species are always more closely related to one another than six species can ever be

Answer on page 31.

Check yourself

The check for this guide is online: a fresh paper of past-paper style questions from the exam bank on D1.3 each time, marked as you go, and saved to your Exam practice progress when you are signed in. Scan the code, or type the link.

biologybybradford.com/exam-maker/sit?topics=D1.3&level=HL&paper=quiz&marks=15&src=ep

Answers

PREDICTIONS IN THE CHAPTERS

01 A gene of 900 bases becomes 901 bases long. Which kind of mutation could this be?

B **An insertion, because one extra base has been added somewhere in the sequence**

An insertion. A substitution swaps a base without changing the length, and a deletion makes the gene shorter. Only an added base makes it longer.

02 The codon GGU codes for glycine, and so do GGC, GGA and GGG. The U is replaced by an A. What is the result?

B **A silent mutation, because GGA still codes for glycine**

Silent. GGA codes for glycine, as GGU did, so the polypeptide is unchanged. The base sequence changed, but whether a mutation is silent depends on the amino acid, not on the bases.

03 A different substitution in the same codon turns GAG into GAA. What would happen to the haemoglobin?

C **Nothing would change, because GAA also codes for glutamic acid, so the chain is the same**

Nothing would change. GAA and GAG both code for glutamic acid, so this substitution is silent. Try it in the tool below: the same codon, two different substitutions, two very different results.

04 Three edits to the same gene: one base substituted, one base deleted, three bases in a row deleted. Which is most likely to stop the protein working?

B **One base deleted, because it shifts the frame, so every codon after it changes**

One base deleted. It shifts the reading frame, so every codon after it changes and a stop codon soon appears. Three bases deleted take out one amino acid and keep the frame; a substitution changes one codon at most. Try all three in the tool in chapter 2.

05 A student says: "A population of insects in a hotter climate will mutate so that it can cope with heat." What is wrong with this?

C **Mutations happen at random with respect to need; heat does not cause the helpful ones**

Mutations happen whether or not they help. Some insects may already carry alleles that cope with heat better; those insects survive and breed more, so the alleles become commoner. That is natural selection acting on random variation, not mutation on demand.

06 A man has a mutation in a lung cell after years of smoking. Could his children inherit it?

A **No, because a lung cell is a somatic cell, and only germ cells make the gametes**

No. The mutation is in one line of lung cells and their descendants. Sperm come from germ cells, which do not carry it, so it cannot be passed on, whatever it does to his lungs.

07 A gene is knocked out in mice and the mice cannot smell. What is the best conclusion?

B **The gene's product is needed for a normal sense of smell in mice**

The gene's product is needed for a normal sense of smell. The experiment does not show that it is the only gene involved, or where else it is used: other knockouts and other tests would be needed for that.

08 Scientists want to change one base in a gene, not knock the gene out. What must they add besides Cas9 and a guide RNA?

B **A template: DNA with the base they want, for the cell to copy as it repairs the cut**

A template. Cas9 only cuts. Left alone, the cell joins the ends and often loses bases, which knocks the gene out. With a template carrying the wanted base, the cell can copy it into the gene.

09 In a conserved region of a gene, the amino acids are the same in ten species but the third bases of the codons differ. Which hypothesis does this support?

B **Function, because silent changes build up while changes to the amino acids do not**

Function. Mutations clearly happen here, since the third bases differ. Only the changes that alter amino acids are missing, which is what selection for a working protein would give.

QUESTIONS ON THE DRY LAB

Q1 In region B the first and second positions hardly ever differ, but third positions often do. Which hypothesis do the data support?

B **Function, because mutations happen there but those that change amino acids are removed**

Function. The third positions show that mutations happen in region B as often as elsewhere. Most third-position changes are silent, so they survive; changes at first and second positions usually change the amino acid, and selection has removed them. The protein needs these amino acids.

Q2 Region C differs little at every kind of position, even at the third positions where region B varies freely. Why can the data not settle, on their own, why it is conserved?

B **Because slower mutation and very strong selection, even on silent changes, would look the same**

Both hypotheses fit. A region that rarely mutates would look like this, but so would one where even silent changes are removed, for example because the DNA itself is a binding site. A further test is needed: measure the mutation rate there directly, or change the sequence and see whether the organism is affected.

Q3 Why does a count with three species usually find fewer differing positions than the same count with six?

A **A position counts as differing if any species differs, and with more species there are more chances for one to differ**

Each species adds its own changes, so more species give more chances for a position to differ. That is why counts with different sets of species scatter, and why a fair comparison between regions uses the same species for each.

How this was made. The course structure, the activities and the teaching choices come from my own classroom, my presentations and my years of teaching IB and AP Biology. I use Claude, an AI, to help write and build the interactive guides, lesson plans, slides and practice questions, and to draw the diagrams and code the interactives. I direct the work and review it against the IB and AP specifications before it goes live.

Written and drawn for Biology by Bradford, biologybybradford.com. Free to print and share with a class.