HELICATE · A-LEVEL BIOLOGY · TOPIC BANK

3.8.3 Using genome projects

Mark scheme

Version 2 · 17 October 2026

TOTAL MARKS	QUESTIONS
37	5

How to use this mark scheme

;	Separates marking points. Each semicolon marks the end of one creditable point, and each point is worth one mark.
;;	Two marks for this answer (used for calculations).
/	Alternative words or phrases that are credited for the same point.
OR	An alternative answer for the same mark.
( )	Words in brackets are not needed to gain the mark.
max	More marking points are listed than marks available. Award up to the number of marks shown.
Accept	An alternative answer that gains credit.
Reject	A wrong answer. It does not gain credit and cancels a correct point it contradicts.
Ignore	Neither credited nor penalised. Mark the rest of the answer.
ECF	Error carried forward: credit correct working that uses an earlier wrong answer.
AO	Assessment objective: AO1 knowledge and understanding; AO2 applying knowledge; AO3 analysing, interpreting and evaluating.

Where an answer is correct but not listed, credit it if it is biologically accurate and answers the question at the depth the marks require. Right + wrong = wrong: a contradiction cancels a correct point.

Not affiliated with or endorsed by AQA.

Genome sequencing projects 3.8.3

Question	Marking guidance	Mark	Comments
01.1	The genomes of a wide range of organisms, including humans, have been sequenced;	**1** (1 × AO1)	Sequencing projects have read the genomes of a wide range of organisms, including humans; methods are continuously updated and have become automated.

Question	Marking guidance	Mark	Comments
01.2	Fruit fly;	**1** (1 × AO2)	Fruit fly: 1.4×10^8 base pairs = 140 million; 14 000 ÷ 140 = 100. Other values: bacterium 929, yeast 500, nematode worm 200.

Question	Marking guidance	Mark	Comments
01.3	Correct answer for 2 marks: 4.65×10^7;; Accept for 1 mark: 46 500 000 OR 46.5×10^6 (correct value not in standard form) **OR** $1.5 \div 100 \times 3.1 \times 10^9$ (correct working)	**2** (2 × AO2)	**Accept** 4.7×10^7 for 2 marks **Reject** 4.65×10^9 (percentage not divided by 100) for any mark

Question	Marking guidance	Mark	Comments
01.4	1. Humans have only about 8 times as many (protein-coding) genes as the bacterium (20 000 ÷ 2 600 = 7.7); 2. Most of the human genome is non-coding (DNA) OR only 1.5% of the human genome codes for proteins, compared with 87% in the bacterium; 3. The human genome has only about 19 times as much coding DNA as the bacterium ($4.65 \times 10^7 \div 2.436 \times 10^6$); 4. So the claim is not supported OR the size of the proteome depends mainly on the number of protein-coding genes/the amount of coding DNA, not the total size of the genome; 5. The first part of the claim is correct ($3.1 \times 10^9 \div 2.8 \times 10^6 = 1107$, more than 1000 times);	**3 max** (2 × AO2) (1 × AO3)	1. **Accept** 7.7 or 7.69 times 1. **Accept** 'the nematode worm has as many protein-coding genes as humans (20 000) but a genome about 31 times smaller, so genome size does not predict gene number' 3. **Accept** any value from 19 to 20 (eg 'about 20 times'); accept ECF from the answer to 01.3 4. Point 4 needs a reason from points 1–3 4. **Accept** 'one gene can code for more than one protein, so the human proteome may be more than 8 times larger, but still nowhere near 1000 times' 5. Point 5 alone does not evaluate the claim; it scores only alongside other points **Ignore** 'humans are more complex' unqualified

Question	Marking guidance	Mark	Comments
01.5	1. Sequencing has become automated/computerised OR sequencing methods have been continuously updated/improved; 2. So many (DNA) fragments/samples/genomes are sequenced at the same time OR machines run continuously without people carrying out each step;	**2** (2 × AO1)	1. **Accept** 'carried out by machines/computers' 2. Must explain why sequencing is faster; repeating that methods are automated/improved does not gain point 2 **Ignore** 'better technology' unqualified **Ignore** named sequencing methods **Ignore** gene machine **Ignore** PCR / electrophoresis / DNA probes

Genomes of simpler organisms and vaccines 3.8.3

Question	Marking guidance	Mark	Comments
02.1	1. Base sequence of the genes is used to determine the amino acid sequences of the proteins OR the proteome is determined; 2. (Proteins identified as) potential antigens (for the vaccine); 3. Only the 580 surface proteins are potential antigens, so the scientists did not need to make and test all 2 160 proteins OR it reduced the number of proteins to test (saving time/cost);	**3** (2 × AO1) (1 × AO2)	1. **Accept** 'identify the proteins (that derive from the genetic code)' / 'the proteins the bacterium can make' 1. **Accept** '(DNA) triplets/codons used to predict the amino acids' 2. **Accept** 'identify antigens'; the term antigen is required 3. **Accept** 'proteins inside the cell cannot be reached by antibodies' only if linked to not needing to make/test these internal proteins 3. **Ignore** restatement of the stem (antibodies bind only to exposed proteins) without why the prediction was useful **Reject** any answer suggesting the vaccine contains antibodies **Ignore** antibiotics **Ignore** 'make medicines/drugs' unqualified

Question	Marking guidance	Mark	Comments
02.2	1. To show that survival/protection is due to the protein (injected) and not due to the salt solution/the injection (itself); 2. To show (how many) mice die/survive without protection OR to show that the dose of bacteria kills most mice (2 of 20 survive);	**2** (2 × AO3)	**Accept** 'as a control' only if the purpose is explained **Ignore** 'to make it a fair test' unqualified **Ignore** 'to compare' unqualified

Question	Marking guidance	Mark	Comments
02.3	800(%);	**1** (1 × AO2)	(18 − 2) ÷ 2 × 100 = 800% **Reject** 900%

Not affiliated with or endorsed by AQA.

Question	Marking guidance	Mark	Comments
02.4	1. Protein R gave the most protection OR the highest number of mice alive (18/20 compared with 2/20 with salt solution only); 2. Only 72% of strains have the gene for protein R, so the vaccine may not protect against the other 28% of strains; 3. Protein T (or P) is found in all strains, and T protected 13/20 mice, so a vaccine containing R and T could protect against more strains; 4. The mice were infected with only one strain, so protection against other strains was not tested; 5. Protein R did not protect every mouse (2 of 20 died); 6. Having the gene does not show that a strain makes the protein OR the protein may have a different amino acid sequence/shape in other strains; 7. Mice may not respond in the same way as humans; 8. Only 20 mice in each group OR no statistical test, so the differences (eg between R and Q) may be due to chance; 9. Mice were infected only 4 weeks after injection OR survival was measured for only 10 days, so long-term protection is not known;	**4 max** (4 × AO3)	**Max 3** if no data from Table 2 are used **Max 3** if only points for or only points against the suggestion are given 2. **Accept** 'more than a quarter of strains do not have the gene for R' 3. **Accept** a named combination including R with T or P, with a reason **Ignore** 'repeat the experiment' unqualified

Question	Marking guidance	Mark	Comments
03.1	1. The base sequence of the gene (for the attachment protein) is found; 2. (Base triplets/genetic code) used to predict the amino acid sequence (of the attachment protein); 3. Different versions have different base sequences (in the gene) OR a change in the base sequence gives a different amino acid sequence;	**2 max** (2 × AO2)	2. **Accept** '(part of) the proteome is determined' **Ignore** references to mRNA unless the base sequence of the gene is used **Reject** 'the amino acid sequence is sequenced directly' for point 2

Question	Marking guidance	Mark	Comments
03.2	1. Version 3, because its percentage has increased every year (from 4% to 61%) OR it was the most common version in 2024 (so it is likely to be the most common in 2025); 2. The attachment protein is the (potential) antigen (on the virus surface); 3. The antigen in the vaccine should be the same as/match the antigen on most of the viruses that will infect people (in 2025); 4. Version 1 is decreasing (from 81% to 18%), so a vaccine with version 1 would match fewer viruses; 5. But version 3 alone would not match about 39% of viruses (versions 1 and 2) OR version 2 has not decreased/has risen slightly (from 15% to about 21%), so it may still be common in 2025;	**3 max** (2 × AO2) (1 × AO3)	1. Version 3 must be linked to a reason from the graph **Accept** readings within ±2% of the plotted values used in points 1, 4 and 5 (4, 61, 81, 18, 15, 21) 5. **Accept** 37–41% for the 39% not matched by version 3 **Reject** any answer suggesting the vaccine contains antibodies

Genomes and proteomes of more complex organisms 3.8.3

Question	Marking guidance	Mark	Comments
04.1	The gene contains non-coding sequences within it;	**1** (1 × AO1)	The non-coding sequences within a gene are introns; they are removed (spliced out of pre-mRNA), so the coding sequence is split. A degenerate code still gives one amino acid for each triplet; histones do not change the base sequence.

Question	Marking guidance	Mark	Comments
04.2	1. Only 1.5% of the human genome codes for proteins, compared with 87% of the bacterial genome OR most of the human genome is non-coding (DNA); 2. So it is harder to find the genes/coding sequences in the human genome; 3. 95% of human (protein-coding) genes contain introns, but (almost) no bacterial genes do; 4. Introns are removed (from pre-mRNA by splicing) OR introns are not translated, so the base sequence of a gene does not directly give the amino acid sequence; 5. The human genome has more regulatory genes (1 600 compared with 140); 6. Regulatory genes control which genes are expressed/transcribed (in a cell/at a time), so with more regulatory genes, the human genome shows less about which proteins are actually made;	**5 max** (3 × AO1) (2 × AO2)	**Max 3** if no data from Table 3 are used 3. **Accept** 'bacteria have no introns' 5. **Accept** a higher proportion of regulatory genes (8% compared with 5.4%) 6. **Accept** 'control transcription/translation of other genes' 6. Must be linked to humans having more regulatory genes (point 5) or more control of gene expression **Ignore** 'switched on/off'; point 6 needs expression or transcription **Ignore** larger genome / more genes / more DNA unqualified **Ignore** histones and membrane-bound organelles

Question	Marking guidance	Mark	Comments
05.1	27.2(%);	**1** (1 × AO2)	1 440 ÷ 5 300 × 100 = 27.170 **Accept** answers from 27 to 27.2

Not affiliated with or endorsed by AQA.

Question	Marking guidance	Mark	Comments
05.2	1. Only some genes are expressed at each stage OR proteins from only 1 120 of the 5 300 genes were detected at the salivary gland stage; 2. Different proteins are made at different stages OR many proteins were found at only one stage (eg 410 only at the salivary gland stage); 3. Regulatory genes control which genes are expressed/transcribed (at each stage); 4. *Plasmodium* is a eukaryote, so its genome contains introns/non-coding DNA, so the amino acid sequences/proteome are harder to predict from the genome; 5. Test proteins present at the salivary gland stage, because these cells enter the blood after the bite (before they reach the liver); 6. Especially the 330 proteins present at both the salivary gland and liver stages, as these are likely to be involved in entering/infecting liver cells; 7. Choose proteins on the (cell) surface (of *Plasmodium*), as these could act as antigens; 8. Proteins not detected may still be present (in very small amounts), so the table may not show every protein made at a stage;	**6 max** (3 × AO1) (1 × AO2) (2 × AO3)	**Max 4** if no suggestion of which proteins to test (points 5–7) **Max 4** if no explanation of why the genome alone is not enough (points 1–4) 6. **Accept** 'the 410 proteins found only at the salivary gland stage' only with point 6's reason: they are made only by the stage that travels from the bite to the liver, so they are likely to be involved in reaching/entering liver cells; 'because they are in the stage that enters the body' is point 5 and cannot score twice **Ignore** 'switched on/off'; point 3 needs expression or transcription **Ignore** larger genome / more genes unqualified **Reject** any answer suggesting the vaccine contains antibodies

Copyright © 2026 Helicate. All rights reserved.