AB4D Genetic Information, Variation and Relationships: Depth and Exam Drill

Total Marks

AQA 7402 · Calculator allowed · about 150 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 This question tests core definitions from molecular genetics.

(a) Give the definition of a gene. **(1)**

(b) Give the definition of an allele. **(1)**

(c) Give the definition of a codon. **(1)**

(Total for Question 1 is 3 marks)

---

2 Structure and replication. A short section of a DNA molecule has the base sequence on one strand: 5 prime - ATG CCT GAA TTA - 3 prime. For the questions below assume standard base pairing and that transcription follows the usual directionality.

(a) Write down the sequence of the complementary strand in 3 prime to 5 prime orientation. **(1)**

(b) State the sequence of the mRNA transcribed from the original 5 prime strand, written 5 prime to 3 prime. **(1)**

(c) Translate the mRNA sequence into the three-letter amino acid code (use standard codon table: AUG = Met, CCU = Pro, GAA = Glu, UUA = Leu). **(2)**

(Total for Question 2 is 4 marks)

3 Hardy-Weinberg and mutation. In a large population of wildflower plants the recessive allele a causes white petals, while the dominant allele A gives purple petals. A sample survey estimates that 9% of the population displays white petals.

(a) Assuming Hardy-Weinberg equilibrium, calculate the frequency of the recessive allele a in the population. Show your working and give your answer to 3 significant figures. **(3)**

(b) Using your answer in (a), calculate the expected percentage of heterozygous plants in the population. Give your answer to 3 significant figures. **(3)**

(c) A mutagen exposure increases the mutation rate from A to a so that each generation an additional 0.5% of A alleles mutate to a. Assuming other Hardy-Weinberg conditions hold and the population is large, explain qualitatively how this will affect allele frequencies over many generations. **(2)**

(Total for Question 3 is 8 marks)

---

4 Chi-squared test on inheritance. A geneticist examines inheritance of a single gene with complete dominance in pea plants. She crosses two heterozygotes and obtains 160 offspring: 118 show the dominant phenotype and 42 the recessive phenotype. Use chi-squared test at 5% significance level with expected ratio 3:1. Critical value for df = 1 is 3.84.

(a) State the null hypothesis for the chi-squared test in this context. **(1)**

(b) Calculate the expected numbers for dominant and recessive phenotypes under the null hypothesis. **(2)**

(c) Compute the chi-squared statistic using chi-squared = sum((observed - expected)^2 / expected) and show your working. Give the value to 2 decimal places. **(4)**

(d) State whether the null hypothesis should be rejected at the 5% level and explain your conclusion using the critical value provided. **(3)**

(Total for Question 4 is 10 marks)

---

5 Practical design and evaluation. A student plans an experiment to measure the rate of incorporation of radiolabelled thymidine into DNA during S phase in cultured mammalian cells. The student will apply identical amounts of radiolabel to cell populations sampled every 30 minutes over 4 hours and measure radioactivity incorporated per microgram of extracted DNA using a scintillation counter. Describe and evaluate the experimental design and suggest improvements, considering controls, sources of error, uncertainty and how data could be analysed. You are assessing the quality of the plan.

(Total for Question 5 is 6 marks)

6 Mutation rate calculation. A study sequences a 2.0 kilobase coding region from 1000 individuals. Over the study period one base substitution that changes a synonymous site is observed in a single individual in this region. Estimate the per base per generation mutation rate assuming each sequenced individual represents an independent germline transmission and that one generation per individual. Express your answer in scientific notation to two significant figures. Show your working.

(a) Calculate the per base per generation mutation rate. Give your answer to two significant figures in scientific notation. **(6)**

(b) Discuss two reasons why this estimate may underestimate the true spontaneous mutation rate. **(2)**

(Total for Question 6 is 8 marks)

---

7 Selection coefficient and change in allele frequency. In a population the beneficial allele B has fitness 1.02 in heterozygotes and 1.04 in homozygotes compared with 1.00 for the alternative allele b homozygote. The current frequency of B is p = 0.10. Approximate the change in p in one generation using the selection coefficient for additive selection (delta p approx = p q s, where q = 1 - p and s is the selection coefficient for the allele). Take s approximated as the advantage of heterozygote over b homozygote for small p. Calculate delta p to 3 significant figures.

(a) Compute q and estimate s from heterozygote fitness, then calculate delta p approx = p q s. Give your answer to 3 significant figures. **(5)**

(b) Explain briefly one limitation of using this simple approximation to predict long-term allele frequency change. **(2)**

(Total for Question 7 is 7 marks)

---

8 Recombination mapping. In Drosophila a geneticist measures recombination between three loci A, B and C on the same chromosome. She finds 120 recombinants between A and B, 30 recombinants between B and C and 150 recombinants between A and C, out of 1000 total progeny. Assume no interference for the simple map estimate.

(a) Calculate the map distance in centiMorgans between A and B, B and C, and A and C. Give each to 1 decimal place. **(3)**

(b) Using these distances, deduce the most likely gene order and explain briefly how the three numbers support that order. **(3)**

(Total for Question 8 is 6 marks)

9 Propagation of uncertainty. A researcher measures two independent variables A and B with results A = 12.3 ± 0.2 arbitrary units and B = 4.56 ± 0.06 arbitrary units (both uncertainties are one standard deviation). She wishes to report the ratio R = A / B and its propagated uncertainty (one standard deviation). Using first-order propagation approximations, calculate R and its uncertainty to two significant figures in the uncertainty and quote R with the uncertainty. Show your working.

(a) Compute R = A / B to appropriate significant figures. **(2)**

(b) Use propagation for division: $(\sigma_R / R)^2 = (\sigma_A / A)^2 + (\sigma_B / B)^2$. Calculate σ_R and report $R \pm \sigma_R$ with σ_R to two significant figures. **(4)**

(Total for Question 9 is 6 marks)

---

10 Molecular phylogenetics and molecular clock. Two species X and Y have mitochondrial gene sequences that differ by 120 nucleotide substitutions over a 1500 base gene. If the neutral substitution rate for this gene is estimated at 1.5×10^{-8} substitutions per site per year per lineage, estimate the time since the two species diverged in millions of years. Show working and give answer to two significant figures. Assume a molecular clock with two independent lineages accumulating substitutions.

(a) Write the equation relating total substitutions $S = 2 \mu t L$ where μ is rate per site per year per lineage, L is sequence length and t is divergence time in years. Rearrange to give t. **(2)**

(b) Calculate t in millions of years using S = 120, $\mu = 1.5 \times 10^{-8}$, L = 1500. Give answer to two significant figures. **(6)**

(Total for Question 10 is 8 marks)

---

11 Synoptic question. Modern agriculture sometimes uses selective breeding to increase crop yield but can reduce genetic diversity. Discuss the potential genetic, ecological and evolutionary consequences of intensive selective breeding in a crop species. In your answer consider genetic variation, inbreeding depression, disease susceptibility, potential for future adaptation, and management strategies to mitigate risks. Support your discussion with appropriate examples and link concepts across genetics, population biology and ecology.

(Total for Question 11 is 8 marks)

Mark scheme · AB4D Genetic Information, Variation and Relationships: Depth and Exam Drill

Question 1

- (a) **B1** a sequence of DNA nucleotides that codes for a polypeptide or functional RNA oe
- **(a) Answer: A sequence of DNA nucleotides that codes for a polypeptide or a functional RNA.**
- (b) **B1** one of two or more alternative forms of a gene at the same locus differing in DNA sequence oe
- **(b) Answer: One of two or more alternative forms of a gene at the same locus that differ in DNA sequence.**
- (c) **B1** a sequence of three nucleotide bases in mRNA that codes for one amino acid or a stop signal oe
- **(c) Answer: A sequence of three nucleotide bases in mRNA that codes for one amino acid or a stop signal.**

Question 2

- (a) **B1** 3 prime - TAC GGA CTT AAT - 5 prime
- **(a) Answer: 3 prime - TAC GGA CTT AAT - 5 prime**
- (b) **B1** 5 prime - AUG CCU GAA UUA - 3 prime
- **(b) Answer: 5 prime - AUG CCU GAA UUA - 3 prime**
- (c) **M1** correct identification of each amino acid from codons
- (c) **A1** Met-Pro-Glu-Leu cao
- **(c) Answer: Met-Pro-Glu-Leu**

Question 3

- (a) **M1** recognise q^2 = frequency of homozygous recessive = 0.09
- (a) **M1** compute $q = \sqrt{0.09} = 0.3$
- (a) **A1** q = 0.300 cao
- **(a) Answer: 0.300**
- (b) **M1** use 2pq with p = 1 - q = 0.700, q = 0.300
- (b) **M1** compute 2 x 0.700 x 0.300 = 0.42
- (b) **A1** 42.0% cao
- **(b) Answer: 42.0%**
- (c) **B1** mutation from A to a will increase the frequency of a each generation, introducing new a alleles to the gene pool
- (c) **B1** over many generations this will shift allele frequencies towards higher q unless balanced by selection or back mutation, leading to more white plants, oe
- **(c) Answer: The directed mutation A -> a will increase the frequency of a generation by generation; over many generations q will rise and the population will have more white plants unless countered by selection or reverse mutation.**

Question 4

- (a) **B1** null hypothesis: observed offspring fit a 3:1 dominant:recessive ratio (differences due to chance)
- **(a) Answer: Null hypothesis: the offspring follow a 3:1 dominant:recessive ratio; any deviation is due to chance.**
- (b) **M1** compute expected dominant = 3/4 x 160 = 120 and recessive = 1/4 x 160 = 40
- (b) **A1** expected dominant 120, expected recessive 40 cao
- **(b) Answer: Expected dominant = 120, expected recessive = 40**
- (c) **M1** substitute values: $(118 - 120)^2/120 + (42 - 40)^2/40$
- (c) **M1** compute each term: (4)/120 = 0.033333... and (4)/40 = 0.1
- (c) **M1** sum terms to give 0.133333...
- (c) **A1** chi-squared = 0.13 cao to 2 d.p.
- **(c) Answer: 0.13**
- (d) **M1** compare chi-squared 0.13 with critical value 3.84
- (d) **M1** 0.13 < 3.84 so do not reject the null hypothesis
- (d) **A1** conclusion: data are consistent with a 3:1 ratio; differences attributed to chance oe
- **(d) Answer: Do not reject the null hypothesis because 0.13 < 3.84; the data are consistent with a 3:1 ratio.**

Question 5

- **Level 3** (5-6): Detailed and balanced evaluation: identifies multiple strengths and weaknesses, suggests realistic improvements to controls, replication and measurement precision, discusses sources of systematic and random error, quantifies uncertainty where appropriate, and outlines an appropriate method of data analysis including plotting, error bars and statistical tests. Clear justification and linkage to experimental aims.
- **Level 2** (3-4): Moderate evaluation: identifies some strengths and weaknesses, suggests at least one sensible improvement (eg replication, a control or calibration), mentions errors qualitatively and suggests a method of analysis such as plotting and use of mean and standard deviation.
- **Level 1** (1-2): Basic comments: mentions one or two obvious issues such as need for controls or replication, limited or no quantitative discussion of uncertainty or data analysis, little justification.
- **Indicative content:**
  - Strengths: time course sampling, normalising to DNA amount, use of scintillation counter provides quantitative readout.
  - Weaknesses: no mention of biological replication, no negative control (no-radiolabel) to measure background, potential variability in DNA extraction efficiency.
  - Improvements: include at least three biological replicates per time point, include blank and positive controls, standardise cell number and extraction protocol, use spike-in control to correct extraction efficiency, calibrate scintillation counter with standards to convert counts to disintegrations per minute, report propagated uncertainty.
  - Sources of error: pipetting error, variable cell cycle synchrony, incomplete removal of unincorporated label, instrument background counts, decay of radiolabel over experiment.
  - Uncertainty and analysis: calculate mean and standard error for each time point, plot incorporation versus time with error bars, fit appropriate model (e.g. linear increase during S phase then plateau), use regression to estimate rate with 95% confidence interval, consider t-test or ANOVA to compare time points if required.
  - Practical notes: ensure radioactive safety, normalise to DNA mass, record detection limits and limits of quantification, express results to appropriate significant figures and propagate errors when dividing counts by DNA mass.

Question 6

- (a) **M1** recognise total number of base observations = 1000 individuals x 2000 bases = 2.0×10^6 bases
- (a) **M1** use mutation rate μ = number of mutations / total bases = $1 / (2.0 \times 10^6)$
- (a) **M1** compute $\mu = 5.0 \times 10^{-7}$
- (a) **M1** convert to scientific notation with two significant figures
- (a) **M1** state $\mu = 5.0 \times 10^{-7}$ per base per generation cao
- (a) **A1** final numerical statement 5.0×10^{-7} cao
- **(a) Answer: 5.0×10^{-7} per base per generation**
- (b) **B1** some mutations may be lethal or strongly deleterious and so not sampled in live individuals
- (b) **B1** sequencing may miss mutations due to coverage limits, filtering of low frequency variants, or errors; somatic mutations not transmitted to germline are not counted
- **(b) Answer: Possible underestimates include: lethal or strongly deleterious mutations that are not present in surviving sampled individuals, and technical factors such as incomplete sequencing coverage, filtering out low frequency variants or inability to detect some mutation types so some true mutations are missed.**

Question 7

- (a) **M1** compute q = 1 - p = 0.90
- (a) **M1** estimate s approx heterozygote advantage = 1.02 - 1.00 = 0.02
- (a) **M1** substitute delta p = p q s = 0.10 x 0.90 x 0.02
- (a) **M1** compute product = 0.0018
- (a) **A1** delta p = 0.00180 cao to 3 s.f.
- **(a) Answer: 0.00180**
- (b) **B1** approximation assumes s constant and small p, neglects dominance deviations and changes in genetic background
- (b) **B1** it ignores genetic drift, epistasis, mutation and changing selection pressures over time, which can alter trajectories, oe
- **(b) Answer: The approximation assumes a constant small selection coefficient and ignores dominance interactions, drift, mutation and environmental changes; thus it may not predict long-term change accurately.**

Question 8

- (a) **M1** compute distances as recombinant proportion x 100
- (a) **M1** A-B = 120/1000 x 100 = 12.0 cM; B-C = 30/1000 x 100 = 3.0 cM
- (a) **A1** A-C = 150/1000 x 100 = 15.0 cM cao
- **(a) Answer: A-B = 12.0 cM; B-C = 3.0 cM; A-C = 15.0 cM**
- (b) **M1** state likely order A - B - C because A-B + B-C = 12.0 + 3.0 = 15.0 which equals A-C
- (b) **M1** explain that sum of adjacent distances equals total distance when no interference
- (b) **A1** conclusion that B is between A and C, so order A-B-C cao
- **(b) Answer: Order A - B - C because A-B (12.0 cM) + B-C (3.0 cM) = 15.0 cM which equals A-C, indicating B lies between A and C.**

Question 9

- (a) **M1** compute R = 12.3 / 4.56
- (a) **A1** R = 2.697 (report 2.70 with appropriate s.f.)
- **(a) Answer: 2.70**
- (b) **M1** substitute fractional uncertainties: $(0.2/12.3)^2 + (0.06/4.56)^2$
- (b) **M1** compute $(0.016260)^2$ approx 2.644e-4 and $(0.013158)^2$ approx 1.732e-4, sum approx 4.376e-4
- (b) **M1** $\sigma_R / R = \sqrt{4.376e\text{-}4} = 0.02093$, so $\sigma_R = R \times 0.02093$
- (b) **A1** $\sigma_R = 2.6974 \times 0.02092 = 0.0564$ -> report uncertainty 0.056 (two significant figures) giving R = 2.70 ± 0.056
- **(b) Answer: R = 2.70 ± 0.056**

Question 10

- (a) **M1** state S = 2 μ L t or equivalent
- (a) **A1** rearrange t = S / (2 μ L) cao
- **(a) Answer: t = S / (2 μ L)**
- (b) **M1** substitute values: $t = 120 / (2 \times 1.5 \times 10^{-8} \times 1500)$
- (b) **M1** compute denominator: $2 \times 1.5 \times 10^{-8} \times 1500 = 4.5 \times 10^{-5}$
- (b) **M1** compute $t = 120 / 4.5 \times 10^{-5} = 2.666... \times 10^{6}$ years
- (b) **M1** convert to millions of years = 2.67×10^{0} Myr
- (b) **M1** round to two significant figures
- (b) **A1** 2.7 million years cao
- **(b) Answer: 2.7 million years**

Question 11

- **M1** discussion of how selective breeding reduces genetic variation by favouring particular alleles and increasing homozygosity
- **M1** explain inbreeding depression: expression of deleterious recessive alleles leading to reduced fitness, seed set or vigour
- **M1** describe increased disease susceptibility due to uniform host genotypes and examples where monoculture led to epidemics
- **M1** discuss reduced adaptive potential: lower standing variation limits response to novel pests, climate change or new selection pressures
- **M1** suggest management strategies: maintain genetic reservoirs, use crop rotation, deploy cultivar mixtures, introgression from wild relatives, and genomic selection to manage diversity
- **M1** evaluate trade-offs: higher yield versus resilience, economic considerations and ethical/social implications
- **M1** link across scales: population genetics processes to ecosystem-level impacts such as effect on pollinators or soil microbiome
- **M1** conclude with balanced judgement and examples (e.g. Irish potato blight, modern wheat breeding strategies) demonstrating synoptic understanding