

Homework 1KRISH PATEL

Q1) i) Probability of getting a nucleotide  $= \frac{1}{4}$ . This is constant for all bases, as they are all  $\frac{1}{4}$  of the independent events.

a) Probability of pulling out AGG  $\rightarrow \left(\frac{1}{4}\right)^3 = \frac{1}{64}$

b) Probability of observing GAA  $= \left(\frac{1}{4}\right)^3 = \frac{1}{64}$

c) Probability of observing A given GAA  $= \frac{1}{4}$   
 (As  $P(A|GAA) = \frac{1}{4} \rightarrow$  As they are independent events (infinite sized bag))

Q2) 8 bases in the bag  $\rightarrow$  No replacement

a) Probability of getting AGG  $= \frac{2}{8} \times \frac{2}{7} \times \frac{1}{6} = \frac{1}{84} \approx 1.2\%$

b) Probability of observing GAT  $= \frac{2}{8} \times \frac{2}{7} \times \frac{2}{6} = \frac{1}{42} \approx 2.4\%$

c) Probability of observing A given GAA  $\rightarrow \frac{0}{5} \rightarrow 0 \rightarrow$  No A's left in the bag.

d) Probability of observing G given GAA  $= \frac{1}{5} = 0.2 \rightarrow 1 \text{ G left in the bag.}$

Q3) a) The error in  $d_1$  could have occurred if there was an incorrect terminator, due to which an additional nucleotide was synthesized (two bases added in the same cycle) thus causing an incorrect read.  $\rightarrow$  Ahead of schedule. AG added instead of just A

b)  $d_1 = AAGTG$

$d_2 = AACTA$

$d_1$  is more likely than  $d_2$ , since an error early on is less likely since it would need to be complementary to the strand. However for  $d_1$ , the G is not a complement to the template strand assuming (length = 5) and thus much likely to get that read (resulting from the unterminated end)