



Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER OF SCIENCE IN BIOMATHEMATICS

END OF SEMESTER EXAMINATION

BMA 8202 - INTRODUCTION TO COMPUTATIONAL STRUCTURAL BIOLOGY

Date: 7th September, 2022

Time: 3 Hours

Instruction: Answer Question one and any other two questions

Question One (30 Marks)

a. Distinguish between the two groups of homologous sequences (3 marks).

b. Classify the following types of mutations as insertions, deletions or substitutions.

ACGA → AGGA

ACGA → ACCGGAGA

ACGGAGA → AGA

(3 marks).

c. Consider the recursive function

Assume n is non-negative

def fib(n):

if n == 0 or n == 1: return 1

else: return fib(n-1) + fib(n-2)

Find fib(5).

(4 marks)

d. What are two main differences between DNA and RNA?

(3 marks)

e. Mention the two important differences between global and local alignment?

(3 marks)

f. Study the following R code

```
f=function(n,y) {
  stopifnot(all(diff(y)>0))
  a = rep(0,n)
  for(c in y){
    for (j in 1:n){
      if(c<=j){
        l=j-c
        w=ifelse(l==0,1,a[1])
        a[j] = a[j] + w
      }
    }
  }
  print(a)
}
a[n]
```

f(7, c(1,2,4))
f(12, c(1,5,10))

- i. What are n and y assumed to be in the function f. (2 marks)
- ii. What is printed and what is the output from the two cases at the bottom of the code? (3 marks)
- g. What are the different architectures used in ANN? (3 marks)
- h. What are Hidden Markov Models (HMMs)? (3 marks)
- i. Suppose that segments of DNA sequences (one chain of the DNA) of Chromosome 1 for 10 individuals are as follow:

```
GTCGAATTGGAATTGG  
GTCGAATTGGAATTGG  
GTCGAATTGGAATTGG  
GCCGAATAGGAATTGG  
GTCGAATTGGAATTGG  
GTCGAATAGGAATTGG  
GCCGAATTTGATTTGG  
GTCGAATTGGAATTGG  
GTCGAATTGGAATTGG  
GTCGAATTGGAATTGG
```

Which positions are SNPs? (3 marks)

Question Two (15 Marks)

- a. Mention the two important differences between global and local alignment? (4 marks)
- b. For the pair of the sequence shown in the Table below, find the optimal alignment score? Use a match score of +1, mismatch score of -2, and gap penalty of -2. (11 marks)

		F	M	D	T	P	L	N	E
	0	-2	-4	-6	-8	-10	-12	-14	-16
F	-2								
K	-4								
H	-6								
M	-8								
E	-10								
D	-12								
P	-14								
L	-16								
E	-18								

Question Three (15 marks)

Consider an experiment of mating rabbits. We watch the evolution of a particular gene that appears in two types, either GG(dominant), Gg(hybrid—the order is irrelevant, so gG is the same as Gg) or gg(recessive). In mating two rabbits, the offspring inherits a gene from each of its parents with equal probability. Thus, if we mate a dominant (GG) with a hybrid (Gg), the offspring is dominant with probability $\frac{1}{2}$ or hybrid with probability $\frac{1}{2}$. Start with a rabbit of given character (GG, Gg, or gg) and mate it with a hybrid. The offspring produced is again mated with a hybrid, and the process is repeated through a number of generations, always mating with a hybrid.

- i. Write down the transition probabilities of the Markov chain thus defined. (4 marks)
- ii. Assume that we start with a hybrid rabbit. Let μ_n be the probability distribution of the character of the rabbit of the n-th generation. In other words, $\mu_n(GG)$, $\mu_n(Gg)$, $\mu_n(gg)$ are the probabilities that the n-th generation rabbit is GG, Gg, or gg, respectively. Compute μ_1 , μ_2 , μ_3 . Can you do the same for general n? (11 marks)

Question Four (15 marks)

- a) Briefly describe the three kinds of RNA molecule that perform different but corporate roles in protein synthesis. (3 marks)
- b) Data of Clarke et al. (1959) reported excess of gastric ulcers in individuals with blood type with blood types A, B, AB, and O as follows:

$$n_A = 186, n_B = 38, n_{AB} = 36, n_O = 284.$$

- i. Write out the likelihood for these data. (4 marks)
- ii. What are complete data categories? (2 marks)
- iii. Express the complete data “counts” as a function of allele frequency estimates and the observed data. (3 marks)
- iv. Apply E-M algorithm to determine the genotype frequencies. (3 marks)