UNIVERSITY OF CAMBRIDGE

MAM2

M. PHIL. IN COMPUTATIONAL BIOLOGY

Friday, 12 May, 2023 2:00 pm to 4:00 pm

COMPUTATIONAL BIOLOGY

Attempt **ALL** questions.

There are **THREE** questions in total.

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
Calculator - students are permitted to bring an approved calculator.

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Scientific Programming

1. Study the following code.

```
m <- function(x, k) {
  j <- 1
  for (i in 1:k) {
    if (x[i] > x[j])
      j <- i
  }
  j
}

f <- function(x, k) {
  i <- 1
  while (i < k) {
    t <- x[i]
    x[i] <- x[k]
    x[k] <- t
    i <- i + 1
    k <- k - 1
  }
  x
}

p <- function(x) {
  n <- length(x)
  for (i in n:1) {
    m <- m(x, i)
    print(m)
    x <- f(x, m)
    x <- f(x, i)
    print(x)
  }
  x
}

p( c(4, 5, 2, 8))                      # case 1
p( c(15, 8, 9, 30, 7, 1, 69, 4, 10))  # case 2
```

What does the code generate from each of the two cases? [7]

What does each function (m, f, p) do? [3]

[QUESTION CONTINUES ON THE NEXT PAGE]

2. Study the following code.

```
q <- function(M, i, j) {

  f <- function(u,v) {
    if (u<1 || u>a) return(0)
    if (v<1 || v>b) return(0)
    return(M[u,v])
  }

  for (i in 1:a) {
    for (j in b:1) {
      s <- f(i-1, j) + f(i, j+1)
      M[i,j] <- M[i,j] * max(s, 1)
    }
  }
  return(M)
}

## case 1
a <- 3; b <- 5
M <- matrix(1, a, b)
q(M, a, b)

## case 2
a <- 4; b <- 5
M <- matrix(1, a, b)
M[2,2] <- M[2,4] <- 0
q(M, a, b)
```

What does the code generate from each of the two cases? [7]

What does the function (q) do? [3]

2 Genomics I

1a Name two major next generation sequencing (NGS) technologies (i.e. not Sanger sequencing). [2]

1b What are their main advantages and disadvantages? [7]

2a Describe the basic structure of a eukaryotic protein-coding gene and the function of the different parts. [10]

2b You have been given the assembled sequence of a novel mammalian genome and asked to annotate it. What are the two main strategies to annotate protein-coding genes? Describe their basic steps. [7]

2c How good are these strategies at predicting the different parts of the basic structure of a eukaryotic protein-coding gene? [5]

3a What are the main types of sequence variants? [2]

3b What are the main types of structural variants? [3]

3c How are SNPs identified? [3]

3d What are the possible types and consequences of SNPs? [6]

3 Biodesign

1. Describe each of the following directed evolution techniques, including an explanation of any key genetic elements, their selection procedures, and an example usecase for each.
   (a) Phage Display [40%]
   (b) Phage-Assisted Continuous Evolution (PACE) [50%]
2. What is negative selection and how might it be applied in Phage Display? [10%]

END OF PAPER