Bioinformatics - Lab 1

Biswas Kumar - Bisku859
Karthikeyan Devarajan - karde799

01/25/2022

Question 1: Hardy- Weinberg equiibrium

We consider a gene locus with two possible alleles (say A and a) and a diploid population with N individuals. Hence, there are 2N alleles in the population. Let p be the proportion of As in the allele population and q the population of as (of course p + q = 1). A population is said to be in Hardy-Weinberg equilibrium if the proportion of AA homozygotes is p^2, aa homozygotes is q^2 and the proportion of heterozygotes (Aa) is 2pq.

Question 1.1

Show that with random mating (i.e. both alleles of the offspring are just randomly, with proportions p and q, drawn from the parental allele population) Hardy-Weinberg equilibrium is attained in the first generation. What is the proportion of A and a alleles in the offspring population? Hence, with random mating, can a population in Hardy-Weinberg equilibrium ever deviate from it?

Solution:

Let P represents the proportion then it is given that:

$$P(A) = p ... eq(1)$$

$$P(a) = q ... eq(2)$$

The population is said to be in Hardy-Weinberg equilibrium if:

$$P(AA) = p^2 ... eq(3)$$

$$P(aa) = q^2 ... eq(4)$$

and

$$P(Aa) = 2pq ... eq(5)$$

Now, the allele proportions at each generation are obtained by pooling together the alleles from each genotype of the same generation according to the expected contributions. Therefore, we can write the equation for P(A) as:

$$P(A) = P(AA) + P(Aa) ... eq(6)$$

and similarly for P(a) we can write it as:

$$P(a) = P(aa) + P(Aa)...eq(7)$$

Further, eq(1), eq(2), eq(3) and eq(4) and utilizing random mating, we can list out proportionas as:

$$P(AA) = P(A) * P(A) = p * p = p^2...eq(8)$$

$$P(aa) = P(a) * P(a) = q * q = q^2...eq(9)$$

$$P(Aa) = P(A) * P(a) = p * q = pq...eq(10)$$

also,

$$P(aA) = P(a) * P(A) = q * p = pq$$

adding equations 8,9 and 10, and putting values from equations 3 and 4, we get:

$$P(AA) + P(aa) + 2P(Aa)$$

which equates to,

$$p^2 + q^2 + 2pq = (p + q)^2 = 1$$

as $p + q = 1$.

and proportion of heterozygotes using equation 10 is:

$$P(A, a) + P(a, A) = 2pq$$

Since it fulfills all conditions of Hardy-Weinberg equilibrium, We can therefore say that with random mating,Hardy-Weinberg equilibrium is attained in the first generation and the population does not deviate from it.

ref : https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle

Question 1.2

We look at the MN blood group has two possible co-dominating (both contribute to heterozygotes) alleles L^M (denoted M) and L^N (denoted N). In a population of 1000 Americans of Caucasian descent the following genotype counts were observed, 357 individuals were MM, 485 were MN and 158 were NN. Use a chi-square goodness of fit test to test if the population is in Hardy-Weinberg equilibrium.

Solution :

```
#Given that
n<-1000   # population
mm<-357   # number of mm
mn<-485   # number of mn
nn<-158   # number of nn
P_mm<-mm/n  #proportion of mm in population
P_mn<-mn/n  #proportion of mm in population
```

```r
P_nn<-nn/n  #proportion of mm in population

#Now calculting p, q, p2 and q2 (i.e all parameters) for Hardy-Weinberg equilibrium

#proportion of N and M in the alleles are represented by p and q respectively :
# one way to calculate
p<-P_mm+(0.5*P_mn)
q<-P_nn+(0.5*P_mn)

#another way to calculate p and q
#total no of alleles is 2n
total<-2*n
p<-(mm*2+mn)/total
q<-(nn*2+mn)/total

#cross check
p+q  # p+q =1
```

```
## [1] 1
```

```r
#Now, calculating p_square, q_square and two_pq

p_square<-p^2
q_square<-q^2
two_pq<-2*p*q



#chi-square goodness of fit test (one method)

chi_test = chisq.test(c(P_mm, P_mn, P_nn), p=c(p_square, two_pq, q_square))
```

```
## Warning in chisq.test(c(P_mm, P_mn, P_nn), p = c(p_square, two_pq, q_square)):
## Chi-squared approximation may be incorrect
```

```r
print(chi_test)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  c(P_mm, P_mn, P_nn)
## X-squared = 9.9938e-05, df = 2, p-value = 1
```

The chi-square goodness of fit test was performed and the obtained p-value is 1. This signifies that we can not reject the hypothesis that the population is in a Hardy-Weinberg equilibrium state.

ref : 1. https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle
2. https://www.biologysimulations.com/post/how-to-use-chi-squared-to-test-for-hardy-weinberg-equilibrium#:~:text=Chi%2Dsquared%20is%20a%20statistical,selection%2C%20and%20large%20population%20size.

Question 2

2.1

Name the protein products of the CDS.
The name of the protein products is RecQ type DNA helicase.
2.2
Write the first four amino acids.
The first four amino acids is MVVA
M - Methionine
V - Valine
A - Alanine
2.3
Save (and submit) the nucleotide sequence of the coding strand that corresponds to these amino acids as a FASTA format file.. Use backtranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/, note species used) to obtain the sequence from the protein sequence.

The reverse sequence for the nucleotide sequence:

```
##     X.c5662.1.Schizosaccharomyces.pombe.chromosome.I..complete.sequence
## 1          GCTACTGGTGGTACTTGTGGTACTTGTGGTTGTACTACTTGTGCTGGTGCTGCTGCTACT
## 2          ACTGGTTGTACTGCTGCTGCTGGTACTTGTGGTTGTACTACTTGTGCTGCTGCTGCTGCT
## 3          TGTACTGGTTGTACTGCTGGTGCTGGTGCTACTGCTACTTGTGGTTGTTGTGGTGGTGCT
## 4          ACTGGTTGTACTACTACTGCTTGTACTACTGGTACTTGTGCTGCTACTGGTACTGGTGGT
## 5          GCTGCTTGTACTTGTGCTGCTACTACTACTGGTGCTACTGCTGCTACTGGTACTGCTGGT
## 6          GCTGCTGCTGGTGGTGCTACTTGTGGTACTACTTGTGCTGCTTGTGCTACTACTACTTGT
## 7          GCTGCTGGTGGTGCTGCTACTGGTACTTGTGGTGCTACTGCTTGTGCTGGTGCTGGTGCT
## 8          TGTGGTGCTGCTGCTTGTACTACTGGTACTGCTGCTGCTGGTGCTTGTGGTGCTACTGGT
## 9          GCTACTGCTACTTGTGGTACTTGTGGTACTTGTACTGCTACTGGTGCTGCTTGTTGTTGT
## 10         GCTGGTTGTACTTGTACTACTACTTGTGGTACTACTTGTGCTGGTGGTGCTTGTGGTGCT
## 11         GGTGCTGCTGCTGCTGCTGGTGGTGCTACTGCTGCTGGTTGTTGTGCTGCTACTTGTGCT
## 12         ACTTGTGGTACTACTGGTGCTGGTGGTTGTACTGGTTGTTGTGCTGGTACTGGTGCTGGT
## 13         GGTTGTACTGCTTGTTGTGCTGGTACTGGTGCTGGTGGTGCTGGTGGTTGTACTACTGGT
## 14         TGTGCTGCTACTACTTGTACTACTTGTTGTGCTGCTGGTGGTGCTGGTTGTGGTACTTGT
## 15         GCTGCTTGTACTGCTTGTTGTTGTGGTTGTACTTGTACTTGTACTTGTACTGGTTGTACT
## 16         TGTACTTGTACTTGTACTGGTTGTACTTGTACTTGTACTTGTACTGCTTGTACTTGTACT
## 17         TGTGCTTGTACTGCTTGTGGTGCTGGTACTGGTTGTGCTGCTGCTACTGGTGCTTGTGGT
## 18         GCTTGTACTACTGGTACTGGTGGTGCTTGTACTGGTTGTGCTTGTGGTGGTTGTACTGCT
## 19         GCTACTACTACTGGTGGTTGTGCTGCTGCTGGTACTGCTTGTGCTGCTGCTACTGGTGCT
## 20         TGTGCTTGTTGTGCTGCTGCTACTACTGGTGGTGCTTGTGCTGCTTGTGCTGGTTGTTGT
## 21         TGTGCTACTTGTGCTACTTGTGCTGCTGCTACTACTGCTACTGCTTGTACTGGTGCTTGT
## 22         ACTACTGGTGCTGCTACTTGTGCTTGTGCTGCTGGTACTACTGGTGGTTGTACTGCTGCT
## 23         ACTACTGCTACTGGTGGTGCTTGTACTTGTACTTGTTGTGCTACTACTACTACTGGTACT
## 24         TGTACTGCTACTACTTGTGCTACTGGTTGTACTACTACTGGTGCTACTGGTACTGGTTGT
## 25         GGTACTGCTGGTGCTGCTACTGGTACTGGTGCTGCTACTGGTTGTACTACTGGTACTACT
## 26         GCTGCTGCTTGTGGTACTGGTGCTACTACTTGTGCTTGTGCTTGTACTGGTTGTGGTTGT
## 27         GCTGCTTGTGCTACTGCTACTGGTTGTGCTGCTGCTACTACTGGTACTGGTTGTGCTACT
## 28         GCTGCTGCTTGTACTGCTGGTGCTGGTACTACTGCTGCTGCTTGTGGTGCTGCTGGTGCT
## 29         ACTTGTACTACTACTACTGGTACTGGTGGTACTACTACTTGTGCTGCTGGTGCTGCTTGT
## 30         ACTGCTGCTGGTGGTGCTTGTACTACTACTGCTGCTGCTGGTACTACTGCTGCTGCTGCT
## 31         GCTGGTACTTGTTGTGCTGCTTGTGCTGCTGCTACTGGTACTACTACTACTGCTTGTGCT
## 32         GCTGCTTGTACTTGTGCTACTACTTGTACTACTTGTACTTGTGCTGCTGCTTGTTGTTGT
## 33         GCTACTGGTACTGCTACTGCTTGTTGTTGTGCTACTGCTTGTGCTACTTGTGCTGGTGCT
## 34         GGTGGTGCTTGTACTGCTTGTTGTGCTGGTACTGCTTGTACTACTACTACTGCTGCTGCT
```

```
## 35          ACTGGTGGTGCTACTGCTACTGGTGCTGGTACTGGTACTGGTACTGGTTGTTGTGGTACT
## 36          GGTTGTGCTTGTGCTGCTGCTGCTGCTTGTGGTGGTGCTGCTTGTGCTGGTGGTGCT
## 37          ACTACTTGTGGTACTGCTTGTGCTACTGGTTGTGCTGCTACTTGTGCTACTGGTGGTGCT
## 38          ACTGCTTGTTGTACTACTACTTGTGGTTGTTGTGCTACTTGTGCTACTGGTACTACTTGT
## 39          GGTGCTTGTGGTACTGCTTGTACTTGTGCTACTGGTGGTGCTGCTGCTGCTGGTACTACT
## 40          GCTACTGCTGCTGCTGCTACTACTGGTGGTGCTGCTGCTGCTACTACTGGTACTGCTACT
## 41          GCTTGTGGTACTTGTGGTGCTGCTTGTGCTGGTTGTACTACTACTGGTTGTGCTGCTGCT
## 42          TGTGGTGGTACTACTGCTGCTGGTGCTGCTTGTGCTGCTGGTACTGCTACTGGTTGTACT
## 43          TGTGCTGCTTGTGGTGCTACTGGTACTTGTGCTGGTACTACTTGTACTACTTGTGCTGCT
## 44          GCTGGTACTTGTGGTGCTACTACTGCTACTGGTACTGCTTGTTGTACTACTACTGGTGCT
## 45          GCTACTGGTGGTGCTGGTGGTGCTGGTGCTGCTGGTGCTGCTGGTGCTGCTGGTGCTGCT
## 46          GGTGCTGGTGGTGCTGGTGGTGCTGGTGGTGGTGGTGGTGCTGGTGGTGCTGGTGCTGCT
## 47          GCTGGTGCTGCTGGTGCTACTGGTTGTGCTTGTGCTGCTGCTGCTTGTGCTACTTGTGCT
## 48          GCTGGTGGTGCTGCTGCTGGTGCTGCTACTGGTGGTACTGGTGGTGCTTGTACTACTACT
## 49          ACTGGTTGTACTACTACTACTTGTGCTGCTGCTGCTACTACTACTGCTACTGGTGGTGCT
## 50          GGTGCTGCTGGTGCTGCTACTTGTGCTGCTTGTGCTGGTTGTGGTGGTTGTGGTGGTGGT
## 51          GCTGGTTGTGCTGGTTGTGCTGCTGGTGCTTGTGCTGCTGCTGGTGGTACTGGTGCTGCT
## 52          GCTGCTACTGCTGCTGCTGCTGCTGCTTGTGGTGCTTGTGCTGCTGGTGCTACTGGTGCT
## 53          ACTGGTACTACTGGTGCTTGTTGTGCTGCTGGTTGTGCTGCTTGTGCTGGTGCTTGTGCT
## 54          GCTTGTGCTGCTTGTGCTTGTACTGCTGCTTGTGCTTGTACTGCTACTACTTGTACTACT
## 55          GGTGCTGCTGGTGCTACTGGTGCTACTGGTGCTGCTGCTGCTGCTGGTGCTACTGCTGCT
## 56          TGTGGTGCTACTGGTGCTGCTGGTGCTGGTGGTGCTGCTGGTGCTGGTGGTGCTGCTGGT
## 57          GCTGGTGCTACTACTGGTACTTGTGCTGCTACTGGTTGTGCTTGTGGTGCTGGTGCTGGT
## 58          GCTGCTGCTGCTGCTACTTGTACTACTACTACTGCTGCTGCTACTTGTGCTGCTTGTGCT
## 59          GCTACTACTTGTGCTGCTACTACTGGTGGTGCTTGTGGTGGTTGTGCTGCTACTACTGGT
## 60          ACTGGTGCTGCTGCTGCTGCTGCTTGTACTACTGGTGGTGCTGGTGCTGCTGCTGCTACT
## 61          ACTGGTGGTGGTGCTACTTGTGCTGGTTGTACTGGTGGTACTACTTGTGGTGCTACTACT
## 62          TGTGGTGCTGCTACTGCTACTGCTTGTGGTGCTGCTACTGGTGGTGCTGCTACTACTGGT
## 63          ACTGCTGCTTGTACTACTACTGGTGGTGCTACTGCTTGTTGTGCTACTACTGGTACTGCT
## 64          GCTGCTTGTTGTGCTGGTACTACTGGTGCTACTACTTGTGGTACTACTGCTACTACTGCT
## 65          TGTACTGCTACTTGTGGTGCTGGTGGTGCTACTACTTGTTGTGGTACTTGTGCTACTTGT
## 66          ACTGCTGCTGGTACTGGTGGTGCTGCTACTGGTGCTTGTTGTGCTACTGGTGGTGGTTGT
## 67          GCTACTGGTGCTGGTGCTTGTGGTGCTGCTACTGGTACTACTACTGCTTGTACTTGTGCT
## 68          GCTGGTGGTGCTGGTGGTGCTGCTGGTTGTACTGCTACTGCTGGTACTGGTTGTGCTTGT
## 69          GCTGCTGGTGCTGGTTGTGGTGCTGGTGGTGCTTGTACTACTACTGGTACTTGTGGTTGT
## 70          ACTACTGCTGGTGCTGGTTGTGCTGCTGCTGCTGGTGGTGCTACTGCTTGTGCTGGTACT
## 71          ACTGGTACTGGTGCTGGTGCTACTGCTACTGGTTGTACTTGTGCTGCTGCTGGTACTGGT
## 72          TGTGCTGGTTGTGCTTGTACTGCTACTGCTTGTTGTACTGCTGCTACTGCTACTACTACT
## 73          TGTACTGGTACTACTGCTTGTGGTACTTGTGGTACTTGTTGTGCTACTTGTGCTGGTTGT
## 74          GGTGGTGCTTGTACTTGTGGTGGTGGTGGTGCTACTGCTGCTGGTGCTGCTGGTGCTTGT
## 75          GCTTGTACTACTGGTGGTGCTGCTGGTTGTGCTGCTACTGGTACTGGTACTGGTGGTGCT
## 76          GGTTGTGCTGCTTGTGCTGGTACTGGTGGTGCTGGTGCTGGTGCTGCTGCTGCTGGTGCT
## 77          GCTGGTGGTACTGGTGGTTGTGCTGGTACTGCTGCTTGTGCTGGTTGTGCTGGTTGTGCT
## 78          GGTACTGCTGCTTGTGCTACTTGTGCTGGTTGTGCTGCTTGTGGTACTTGTGGTTGTTGT
## 79          GCTGCTTGTACTACTACTGGTGCTACTGCTGGTACTGGTTGTACTGGTGCTGCTGGTGCT
## 80          TGTGGTGCTTGTGCTGCTACTGGTGCTTGTGCTGCTACTGGTGCTTGTGCTGCTACTGGT
## 81          GCTTGTGCTGCTACTGGTGCTTGTGCTGGTGCTGGTGCTTGTGCTGGTACTGCTGCTTGT
## 82          GCTGCTACTGCTGCTTGTGCTGCTACTGCTGCTTGTGCTGCTACTGCTGCTTGTGCTGCT
## 83          TGTGCTTGTTGTGCTGCTACTGCTTGTACTGGTGCTTGTGGTGCTACTGGTGCTACTGGT
## 84          GCTACTGCTGCTGGTTGTACTGGTGGTTGTGCTACTGCTTGTACTACTGGTGGTGCTGCT
## 85          TGTACTGGTTGTGCTTGTGGTGCTGGTGGTTGTGCTACTACTGCTGCTGCTGCTACTACT
## 86          GGTGGTTGTACTACTACTACTTGTACTGGTTGTGCTGCTTGTGCTGCTACTGCTTGTGGT
## 87          GCTACTACTACTACTACTTGTACTGCTGCTGCTGCTGCTACTGGTACTACTTGTGCTGCT
## 88          GGTGCTTGTACTACTGGTGGTGCTGCTGCTACTTGTGCTACTGGTGGTGCTGGTACTACT
```

```
## 89          ACTTGTACTACTGGTTGTGCTACTGGTTGTGCTACTGGTACTTGTACTACTACTGGTTGT
## 90          GCTACTGCTGCTGCTGGTGCTACTGGTGGTACTGCTTGTACTACTTGTGCTGCTGCTGCT
## 91          ACTGCTACTGGTTGTTGTACTGCTACTGGTGCTGCTGCTACTTGTACTTGTACTGGTTGT
## 92          TGTACTGGTTGTACTACTACTGGTTGTGCTTGTTGTTGTTGTACTGCTGCTACTGCTACT
## 93          GCTTGTGCTTGTGGTACTGGTACTTGTGGTACTTGTACTACTGGTACTGCTGGTTGTGCT
## 94          GGTTGTGCTACTGGTACTGGTGCTGGTACTACTGCTTGTGCTGCTTGTGGTGCTACTACT
## 95          GGTGCTACTTGTGGTGCTACTGGTGCTGCTGCTGCTGCTTGTGCTGCTGCTACTTGTGGT
## 96          GCTTGTACTACTGGTACTACTGGTACTTGTTGTGCTACTACTTGTTGTGCTACTTGTGGT
## 97          ACTACTACTTGTGCTGCTGCTTGTTGTGGTTGTACTGGTGGTACTACTTGTACTGCTACT
## 98          ACTGGTTGTACTACTGCTACTGGTTGTACTTGTGCTACTGGTACTGGTACTACTACTACT
## 99          GGTTGTACTACTACTGCTACTTGTGCTTGTGGTTGTACTACTGGTGGTACTTGTGCTGCT
## 100         GCTGGTGCTGCTGCTACTACTACTGCTACTGCTACTGGTGCTACTGGTACTACTTGTACT
## 101         ACTACTGCTTGTGGTGCTGCTGCTTGTTGTTGTGCTGCTGCTGCTGCTGGTACTTGTGGT
## 102         ACTACTGCTGGTGGTGGTGCTACTGCTACTTGTGCTACTACTTGTGGTGCTGCTTGTGGT
## 103         GGTGCTGCTGGTGGTGGTACTGCTACTGGTTGTGCTGCTGCTACTGCTTGTGCTTGTACT
## 104         GGTTGTGCTGCTGGTGGTGCTTGTACTGCTGCTGGTTGTTGTTGTGCTGCTGGTACTGCT
## 105         TGTTGTGGTACTGCTACTACTGGTACTACTACTTGTGCTGGTTGTTGTGCTTGTGGTGCT
## 106         GCTGGTACTGCTGCTACTACTTGTACTGCTACTGGTACTGCTTGTTGTTGTGGTACTGGT
## 107         TGTGCTACTACTGGTGGTGCTGGTGCTACTGGTTGTGCTACTACTTGTGCTGCTACTGCT
## 108         GCTTGTGCTACTGGTGGTACTGGTTGTGGTGCTACTACTGGTGGTGCTTGTACTACTGGT
## 109         GCTGGTACTGGTGCTGGTTGTACTGGTGCTTGTGCTGGTTGTACTTGTACTGGTACTGCT
## 110         ACTGGTGCTGGTGGTGGTGGTGCTACTGGTACTACTACTGGTTGTTGTGCTGCTGCTGGT
## 111         ACTACTTGTGCTGGTGGTGCTTGTACTACTGGTACTACTGCTGCTGCTGCTGGTGCTGGT
## 112         TGTACTGCTACTGGTACTACTACTACTGGTGCTACTGCTACTGGTGCTGCTACTGGTACT
## 113         GGTGGTGCTGCTGCTGCTGCTTGTACTGCTACTACTGCTTGTTGTTGTGCTACTACTACT
## 114         TGTGGTACTACTGCTTGTACTGCTGCTGGTGGTACTTGTGCTGCTACTACTGGTGGTGCT
## 115         GGTGCTACTGGTGCTTGTGCTACTGCTGCTGCTTGTGCTGCTACTGCTGGTACTGCTGCT
## 116         GCTACTACTGGTGGTGGTGCTACTGCTACTACTTGTGCTACTACTTGTACTACTACTGCT
## 117         GCTGCTGGTGCTGGTACTTGTGGTGCTACTGCTGGTGCTGGTGCTACTGCTTGTGGTACT
## 118         ACTTGTGCTACTTGTACTTGTGCTACTGCTGGTACTGGTACTGGTTGTACTGGTACTACT
## 119         GCTTGTGGTGCTGCTTGTGCTGCTACTGCTTGTACTACTGCTGCTGGTGCTGCTTGTGCT
## 120         GGTACTGGTGCTGCTTGTACTACTACTGGTACTTGTGCTACTTGTGGTGGTACTACTACT
## 121         ACTACTTGTTGTTGTGCTACTTGTGGTGCTACTGGTACTTGTACTGCTGCTGGTGCTGCT
## 122         GGTGGTGCTACTTGTACTGGTGCTTGTGCTGCTGCTGGTACTACTGGTACTACTACTGGT
## 123         GGTACTGGTGGTGCTGGTACTGCTGCTGGTACTGGTGCTACTTGTGCTGGTTGTGCTGCT
## 124         GCTGGTGCTGCTGCTACTGGTGCTGGTACTGGTTGTGGTGCTTGTGCTGCTTGTACTGCT
## 125         ACTGCTGGTTGTGCTGCTTGTTGTGCTTGTACTGCTTGTGCTGCTTGTGGTGCTTGTGCT
## 126         GCTTGTACTTGTACTGCTGCTACTGGTGCTACTGCTGCTACTGGTGCTTGTGCTGCTACT
## 127         GGTGCTACTGGTACTGGTACTACTACTTGTACTTGTGCTGCTGCTACTACTGCTTGTGCT
## 128         ACTACTGGTGGTACTTGTACTGCTGCTGCTACTTGTACTGGTTGTACTGCTACTACTGCT
## 129         GCTGCTGCTGCTGGTACTGCTACTGGTGCTGGTGCTTGTGCTGCTGCTGGTGGTTGTGCT
## 130         ACTTGTACTGCTACTTGTACTACTTGTGCTGCTACTGGTGCTGGTACTACTGCTACTACT
## 131         GGTACTACTACTACTGGTACTTGTACTGCTGGTACTGGTACTGCTTGTGCTACTGCTACT
## 132         TGTACTGGTTGTACTGGTGGTGCTTGTGCTGCTTGTTGTGCTGGTTGTTGTGCTGGTGCT
## 133         GGTTGTGCTTGTGCTGCTGGTGCTGGTGCTACTGGTGGTACTGGTACTGCTACTACTGGT
## 134         GGTGCTTGTACTACTACTGGTTGTGGTGGTGCTGCTACTGGTGGTTGTGCTGCTGGTACT
## 135         GCTACTGCTGCTGGTGCTTGTACTTGTGGTTGTGGTGCTGCTACTACTGGTACTGCTACT
## 136         ACTACTGGTGCTACTGGTACTACTACTGGTGGTGCTGCTGGTGGTTGTACTGGTGCTACT
## 137         GGTGCTACTACTACTGCTTGTGCTGGTTGTGCTGGTGCTACTGCTTGTGGTGCTACTGCT
## 138         GCTGGTGCTTGTACTTGTGGTACTGCTGCTACTGCTACTGGTGCTGCTGGTACTACTACT
## 139         GGTTGTACTGGTGCTGCTGCTGCTGGTTGTTGTGCTGCTACTTGTTGTTGTTGTGCTGGT
## 140         GGTACTACTACTTGTACTACTACTTGTACTGGTGCTGGTTGTTGTGGTTGTACTACTACT
## 141         TGTTGTGCTACTACTACTACTGCTGGTTGTGCTTGTACTACTTGTGGTGGTACTGCTTGT
## 142         ACTGCTACTGGTACTACTACTACTGGTGGTACTACTTGTGGTGCTTGTTGTGCTACTACT
```

```
## 143        GGTGGTGCTGCTGGTTGTGCTACTACTGGTGCTACTGGTGCTGCTGGTACTGCTACTGGT
## 144        ACTGGTGCTTGTGCTGCTTGTTGTGGTTGTACTGGTGCTACTGCTGGTGGTACTTGTGGT
## 145        GCTGCTGCTGGTACTGCTGGTTGTACTGGTACTGCTACTGCTTGTACTACTGGTGGTGCT
## 146        ACTACTACTTGTGCTACTGGTACTACTACTGGTACTGGTGCTACTACTGGTTGTACTGGT
## 147        GGTTGTGGTGCTGCTTGTGGTGCTACTACTGGTTGTGCTGCTGCTGGTGCTGGTGCTACT
## 148        ACTACTGCTTGTTGTGGTACTGCTACTTGTGGTGCTGCTACTACTACTACTACTTGTTGT
## 149        GCTGCTGCTGGTGGTTGTTGTGCTTGTTGTACTGCTTGTTGTGCTGCTACTGGTTGTGCT
## 150        ACTACTTGTGCTGCTGCTGCTGCTTGTTGTGGTACTACTGGTGGTGGTGCTACTACTACT
## 151        TGTGGTGCTGCTGCTTGTACTGCTTGTGCTGGTGGTTGTGCTTGTGCTACTACTGGTTGT
## 152        ACTTGTGCTTGTACTGCTACTACTACTACTGCTGCTGCTGGTGCTGCTGCTGCTGCTGCT
## 153        GCTTGTGCTACTTGTGGTGCTGGTGGTGCTGCTGGTGCTGCTGCTACTGGTGCTTGTGGT
## 154        GCTGGTGGTGGTGCTGCTACTTGTGCTACTGCTACTACTACTTGTGGTGCTACTACTACT
## 155        GCTTGTGCTGGTGGTTGTACTGGTGGTGCTTGTGCTACTGCTTGTGCTTGTGGTGCTGCT
## 156        GCTTGTGCTTGTGCTGGTTGTGGTTGTACTTGTACTGCTTGTGCTACTTGTACTGCTACT
## 157        GGTGGTGGTTGTGGTTGTGCTTGTACTGCTACTGGTGGTGCTTGTGCTGCTTGTACTACT
## 158        GGTTGTGCTACTACTGCTACTTGTACTGGTTGTTGTGGTACTTGTGGTGGTGCTACTACT
## 159        GCTACTACTACTTGTGGTTGTTGTGCTGCTTGTACTACTACTACTACTACTTGTGGTACT
## 160        GGTTGTGCTGCTGGTTGTACTGCTACTGCTGCTGGTACTGGTGGTTGTGCTGGTGGTGCT
## 161        GCTTGTACTGCTACTACTGCTTGTGCTGGTGCTACTACTTGTGGTGCTGGTGCTACTGCT
## 162        GCTTGTTGTTGTGGTGCTTGTTGTTGTGCTACTGGTGGTGCTTGTACTGGTACTACTGGT
## 163        GGTACTGGTGGTGCTGCTGCTTGTGCTGCTGCTGGTTGTGCTTGTTGTTGTGCTACTACT
## 164        TGTGCTACTTGTGCTGCTGGTTGTGGTGCTGGTACTACTGGTGCTACTTGTGCTGCTACT
## 165        ACTGGTGGTGCTGGTGGTGCTGCTGGTTGTGGTTGTACTGGTGCTGCTACTGGTGCTGGT
## 166        GCTGCTGGTTGTACTGGTGGTTGTGCTGCTGGTGGTACTACTGGTGGTACTGCTGGTGGT
## 167        ACTGGTGCTGCTTGTGCTGCTGCTACTGGTGGTACTGGTGGTGCTGGTGGTGGTGGTGGT
## 168        GCTTGTGCTGCTGGTGGTGCTGCTGCTGCTGCTGGTGGTGCTTGTGCTGCTGGTGCTTGTGCT
## 169        GCTGCTACTGGTGCTGGTGGTGCTGGTGCTGCTGGTGCTGCTTGTGCTGCTGGTGGTGCT
## 170        TGTGGTGCTGGTGGTACTGCTGCTGCTGGTGGTTGTACTGGTGCTGCTGCTACTGGTGCT
## 171        TGTGCTTGTGCTGGTTGTTGTTGTGGTACTACTGGTACTGCTGCTGCTACTTGTGCTGCT
## 172        GGTGCTTGTACTTGTACTTGTGCTACTGGTGCTTGTACTACTGCTTGTGCTGCTGGTGCT
## 173        TGTTGTGCTGCTACTACTGGTGGTTGTTGTGCTTGTACTGCTTGTGGTTGTTGTTGTGCT
## 174        TGTTGTGGTTGTGGTTGTTGTTGTGCTTGTTGTGGTTGTGCTACTACTACTTGTGCTTGT
## 175        ACTGCTTGTTGTGGTTGTTGTTGTGCTGGTGGTGCTTGTACTACTTGTACTACTTGTGCT
## 176        GCTTGTTGTGCTACTTGTACTTGTGCTGCTGCTTGTACTACTTGTACTGGTACTACTTGT
## 177        GCTGCTTGTGCTACTACTGGTACTACTGGTACTACTGGTGGTGGTTGTGCTACTACTGGT
## 178        ACTTGTACTTGTGCTGCTACTGCTACTACTGCTACTGGTGGTACTTGTACTTGTGGTGCT
## 179        GCTGGTTGTGGTGCTGCTGCTACTACTTGTTGTGGTACTACTTGTGGTTGTACTGGTGCT
## 180        GCTGCTTGTGCTGCTACTACTACTTGTGCTGCTACTTGTGGTGGTACTACTACTGCTACT
## 181        ACTACTACTACTTGTACTTGTACTACTACTACTGCTGCTGCTACTTGTGGTACTGCTACT
## 182        GGTGCTGCTACTACTACTGGTGCTACTACTGCTTGTTGTGGTACTGCTTGTACTTGTTGT
## 183        TGTGCTGCTTGTACTGGTGGTGCTGGTGGTACTGGTGGTGCTGCTGCTGGTACTTGTACT
## 184        ACTACTGGTACTTGTGGTACTACTACTACTACTGGTGCTACTGCTTGTTGTGGTGGTTGT
## 185        GGTTGTACTTGTGCTACTTGTGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGGTGCTTGT
## 186        GCTGCTGCTTGTTGTTGTTGTGCTGGTGGTGCTGCTGCTGGTGGTACTGGTGCTACTGGT
## 187        GCTGCTACTGCTACTGGTGGTACTTGTGCTTGTGGTTGTACTGGTGGTACTACTTGTACT
## 188        GCTGGTACTGGTTGTTGTTGTGCTACTGGTGCTACTGGTACTTGTGGTACTACTGCTTGT
## 189        GGTGGTTGTGCTGCTGGTGCTACTGCTACTGGTGCTACTGGTTGTACTTGTGCTGGTGCT
## 190        GGTACTGGTGCTGCTACTGGTGCTGCTGCTGCTGGTGGTGGTGCTTGTACTGGTTGTACT
## 191        ACTGGTACTACTACTGGTACTACTTGTGGTGGTGGTGGTGCTGCTACTACTGGTGGTGCT
## 192        TGTTGTGGTTGTGCTACTACTTGTGCTGCTGCTGGTGCTACTGGTACTGCTTGTGGTGCT
## 193        ACTACTGGTGCTTGTACTACTACTGCTGGTGCTGGTGCTTGTGGTTGTGCTGCTTGTACT
## 194        ACTTGTTGTTGTGGTGCTACTACTACTGGTACTACTACTGCTACTTGTACTACTGGTGCT
## 195        TGTGCTACTGCTTGTGGTGCTGGTACTTGTGCTGGTTGTGCTACTACTGCTGCTTGTTGT
## 196        GCTGCTTGTGCTGGTACTGGTGGTACTTGTACTACTTGTGGTGGTACTACTACTACTACT
```

```
## 197        ACTGGTGCTGCTGCTGGTACTACTACTGGTGGTTGTGCTGCTTGTGCTTGTACTACTGGT
## 198        GGTACTTGTGGTACTACTACTGGTGGTTGTGCTTGTGGTGCTGGTACTGGTGGTACTGCT
## 199        GCTACTACTGGTGCTACTGGTGCTGCTGGTTGTGCTTGTGCTACTACTACTGGTACTACT
## 200        GCTTGTACTTGTGCTTGTTGTGCTGGTACTGGTGGTGCTGGTTGTGCTACTGGTGGTGCT
## 201        GGTGCTGCTTGTGGTGGTTGTACTACTACTGGTACTTGTGGTGCTGGTGCTGGTTGTGCT
## 202        ACTTGTGGTTGTGGTGGTACTACTGGTACTTGTACTGGTGGTTGTACTACTGGTACTGCT
## 203        ACTGGTTGTGCTTGTTGTGCTACTACTGGTTGTGCTTGTACTACTGGTACTACTGGTGCT
## 204        GGTACTGGTTGTTGTGCTTGTTGTACTACTACTTGTTGTTGTGCTGGTGGTTGTGCTGCT
## 205        TGTACTGCTGGTGCTGGTGCTACTGGTGGTACTACTGGTTGTTGTGCTGGTGCTTGTGCT
## 206        GCTGCTTGTGGTACTACTACTACTGGTACTGCTTGTGCTGCTGCTTGTACTACTACTACT
## 207        GCTTGTGGTACTACTACTACTGCTTGTGGTGCTGGTGCTGCTGCTTGTGGTACTTGTACT
## 208        GCTTGTACTGGTTGTGCTTGTGGTGGTGGTGCTGCTGCTGCTTGTGCTACTTGTACTACT
## 209        ACTACTGCTTGTACTACTACTACTACTGGTTGTGCTACTTGTTGTTGTACTGCTACTGGT
## 210        GCTACTGCTGCTACTGCTTGTTGTGGTGCTGGTACTACTACTACTACTGGTACTACTGGT
## 211        GGTGCTTGTACTACTGGTTGTGGTGCTGCTTGTGGTACTACTGGTGCTACTGGTGCTGCT
## 212        GCTTGTGGTGCTGCTTGTGCTGCTGCTGGTGGTACTTGTACTACTACTGGTGCTGCTGGT
## 213        GGTACTGGTGCTACTGGTGGTGCTTGTGGTACTGGTTGTACTGCTACTTGTGCTACTTGT
## 214        ACTACTACTACTGGTACTTGTGGTGCTGCTTGTTGTGCTGCTGCTGCTGCTGGTGGTGCT
## 215        ACTGGTACTTGTGGTGCTGCTACTGCTTGTGCTACTACTTGTGCTACTTGTGGTACTTGT
## 216        GGTACTTGTACTACTTGTGCTTGTTGTGCTGGTACTTGTGGTGGTGCTTGTACTACTGGT
## 217        ACTACTTGTGGTTGTACTTGTGCTTGTGCTTGTTGTTGTGCTACTGGTACTGCTGCTTGT
## 218        TGTGCTACTACTACTGCTTGTGCTTGTGCTGGTGGTGCTGGTGCTACTGGTACTGCTGCT
## 219        GGTTGTGGTGCTTGTGGTGCTGCTGGTGCTGCTTGTGGTGCTTGTGCTGCTGCTACTGGT
## 220        GCTGCTTGTACTACTACTGGTGCTTGTGGTTGTGGTACTACTACTTGTGGTGCTGCTGCT
## 221        ACTGGTTGTGCTGCTGCTACTGGTGGTGGTGCTGCTGGTGCTTGTGCTTGTGGTGCTGCT
## 222        ACTTGTGCTACTGGTGCTACTTGTGGTTGTACTGCTTGTTGTGCTGCTGGTGGTTGTGCT
## 223        ACTACTTGTGGTGGTGCTTGTACTTGTGGTGGTACTGCTACTTGTGCTGCTTGTACTGCT
## 224        ACTGCTACTGGTGGTGGTGCTGGTACTGGTTGTGGTACTACTACTGCTGGTACTGCTGGT
## 225        ACTGCTTGTGCTTGTACTGCTACTGGTGGTGCTACTACTGCTTGTTGTGCTGGTTGTACT
## 226        ACTTGTGCTACTTGTACTGCTACTGGTGGTGCTACTACTGCTACTGGTACTGCTTGTGCT
## 227        GGTGGTGCTGGTGCTTGTGCTGGTGGTACTTGTGGTGCTGGTTGTACTGGTGGTGCTGCT
## 228        GGTGCTGGTGCTACTGGTGGTTGTGCTGCTGGTACTGCTACTGGTTGTGGTGCTACTACT
## 229        GGTTGTGCTGGTTGTGCTACTACTGGTACTACTACTACTGCTTGTGGTGCTGGTGCTGCT
## 230        GCTACTGCTACTGGTGCTACTACTTGTACTGCTTGTGCTACTGGTGGTACTTGTGGTGCT
## 231        GGTTGTACTGCTTGTGGTACTGCTGGTGCTGGTGGTGCTACTACTTGTGGTGCTACTGGT
## 232        GCTGCTGCTGCTGCTTGTACTACTACTTGTACTACTGCTGCTACTGGTGCTACTGCTGCT
## 233        ACTGCTTGTGGTGCTACTGGTACTGGTACTGGTACTACTTGTGGTGCTACTTGTGGTACT
## 234        ACTACTTGTACTTGTGGTTGTGCTGCTGGTACTGGTGCTGCTGCTACTGGTGGTGCTACT
## 235        GGTGGTTGTGGTGCTGCTACTGGTACTGGTACTGCTACTGGTACTACTGGTACTGGTTGT
## 236        GCTACTTGTGGTACTACTACTGGTTGTACTGCTGCTTGTACTGGTACTGGTACTACTACT
## 237        GCTTGTACTGGTTGTACTTGTGCTGCTGGTGCTACTGGTTGTACTTGTGCTGGTGCTACT
## 238        ACTTGTGGTACTACTGCTTGTACTACTGGTGGTACTGGTGCTGCTGGTGCTGCTACTTGT
## 239        GCTGCTTGTACTGGTACTGGTACTTGTACTGCTTGTGGTGCTACTGGTACTGCTACTGGT
## 240        GGTGCTGGTACTGGTGCTGCTGCTTGTTGTGGTGCTTGTGCTACTACTGGTTGTTGTGCT
## 241        GGTGCTGCTGCTTGTGCTTGTTGTGGTGCTGCTGCTTGTTGTGCTGGTTGTTGTGCTACT
## 242        ACTGGTTGTGCTGCTTGTGCTTGTGCTACTACTTGTGGTTGTGGTACTACTGCTACTGCT
## 243        GCTACTGGTTGTGCTACTTGTGGTACTACTACTACTTGTGGTACTTGTACTACTTGTTGT
## 244        TGTTGTTGTTGTTGTGCTTGTTGTGCTTGTGCTGGTTGTTGTGCTGGTGGTGGTGCTGCT
## 245        ACTGCTGGTTGTGCTGGTACTGGTGGTACTGCTACTGGTGCTGGTACTGGTTGTACTGCT
## 246        ACTGGTGCTGCTTGTGCTTGTACTGCTGCTTGTGCTTGTACTGCTTGTACTGCTGGTACT
## 247        GCTTGTACTGCTTGTGGTTGTTGTGCTGGTACTGGTACTTGTACTACTACTGGTACTTGT
## 248        GGTGGTGCTGCTACTACTGCTACTTGTGGTGGTGCTGCTGCTACTTGTGCTTGTACTACT
## 249        ACTGGTACTACTTGTTGTTGTACTACTTGTGGTACTTGTACTGGTACTGCTACTTGTGGT
## 250        TGTTGTGCTGCTTGTGCTACTGGTGGTGCTGCTGGTGCTGCTGCTGCTGGTTGTACTACT
```

```
## 251        ACTGGTGGTTGTGCTGCTACTGGTTGTACTGCTGCTACTGCTTGTGGTGCTGCTACTTGT
## 252        ACTGCTGCTGCTGGTACTGCTACTGGTGGTACTACTACTGGTGGTGCTGCTGGTGCTTGT
## 253        GCTACTGGTACTTGTGGTTGTACTTGTGCTGGTTGTTGTGCTACTTGTGGTACTTGTGGT
## 254        ACTGGTGGTACTTGTGCTACTGCTGCTGCTTGTGGTTGTGCTTGTGCTACTGCTTGTGGT
## 255        GCTACTGGTGCTGCTTGTGCTACTACTACTGGTGCTGCTTGTGCTGCTACTGGTACTGCT
## 256        TGTGCTGCTTGTGCTGGTGGTGGTTGTGGTACTTGTGCTGCTTGTTGTGCTACTGGTGCT
## 257        TGTGCTACTGGTGCTGCTACTTGTGGTACTGGTACTACTTGTGCTACTGGTGGTACTGCT
## 258        GGTACTGGTACTACTGGTGGTGCTGGTGGTACTGCTACTGGTGCTGGTACTGGTGGTACT
## 259        GCTACTACTGGTACTACTGGTGGTACTGCTACTACTGGTGGTACTGCTACTACTGGTGGT
## 260        ACTGCTACTACTGGTGGTACTGGTGCTACTGGTGGTACTGGTGCTACTGGTGGTACTGGT
## 261        GCTACTGGTGGTACTGGTGCTACTGGTACTACTGGTGCTACTACTTGTACTTGTGGTACT
## 262        GCTTGTTGTGCTACTACTTGTGCTACTACTACTTGTGGTTGTTGTGGTGCTGGTACTGCT
## 263        ACTGCTGCTGCTGCTGGTACTTGTGGTACTGGTACTGGTTGTGCTGCTGGTTGTTGTGGT
## 264        ACTTGTGCTGCTGCTGCTGCTGCTTGTGCTGCTACTGGTGGTGGTACTGCTGGTGCTTGT
## 265        ACTTGTGCTGCTTGTACTGGTGCTTGTGCTACTACTACTTGTGCTGGTTGTGCTTGTGCT
## 266        GCTTGTACTTGTGGTGCTGGTTGTGGTGCTACTACTTGTACTACTACTTGTGGTGCTGGT
## 267        ACTGGTACTGCTACTGCTGCTGCTGGTGCTACTGGTGCTGCTACTGGTACTTGTACTGCT
## 268        GCTGGTACTACTGGTTGTGCTTGTACTTGTACTACTGGTGGTACTGCTGCTTGTTGTTGT
## 269        GCTGGTGCTTGTGCTTGTGCTGGTGCTGCTGCTACTACTTGTGGTACTGGTTGTACTTGT
## 270        GCTACTGCTTGTTGTGGTGGTGGTGCTGCTGCTGGTTGTGCTACTGGTTGTTGTTGTTGT
## 271        GGTACTGCTTGTGGTTGTACTACTGCTACTTGTACTGCTTGTACTACTGGTACTACTGCT
## 272        ACTGCTGCTGCTACTGGTACTGGTGGTTGTGCTGCTGGTGGTTGTTGTGGTGCTTGTTGT
## 273        GCTTGTGCTGCTTGTACTACTGGTTGTGGTGCTGGTGCTGCTACTGGTACTGCTGCTGCT
## 274        TGTACTGCTTGTGGTACTGCTACTACTTGTGGTGCTACTACTACTTGTGCTGCTGGTGGT
## 275        GCTTGTACTACTACTGGTACTACTACTGCTACTACTACTACTGGTACTGGTGGTACTACT
## 276        ACTGGTGCTTGTGCTGCTGCTGGTACTACTACTGGTGCTGCTTGTGCTACTGGTTGTTGT
## 277        GGTGCTTGTGCTGGTACTGGTGCTACTGCTACTGGTGGTTGTACTACTGCTTGTGCTTGT
## 278        ACTACTTGTACTGGTGCTTGTACTGGTACTGCTGGTGGTACTTGTGCTACTGGTGGTGGT
## 279        TGTACTGCTGGTGCTGCTGCTGGTGGTTGTGCTGCTGCTTGTACTACTGGTGCTACTGCT
## 280        ACTTGTGGTACTACTGCTGGTACTACTACTGCTTGTACTGCTACTGGTTGTACTACTGGT
## 281        GGTGCTGCTTGTGCTGCTACTGGTACTGCTTGTGCTGCTACTGCTTGTTGTGGTGCTGCT
## 282        GGTGGTGCTTGTACTGCTACTACTGGTTGTGCTGGTGCTACTGCTGCTGCTACTACTTGT
## 283        ACTACTGCTTGTGCTGCTGGTGGTACTGGTGCTACTGGTACTGCTTGTGGTACTGGTGCT
## 284                                                                ACTTGT
```

2.4

Compare your obtained coding strand sequence with the nucleotide sequence provided (when following the CDS link). Are they the same or do they differ? Try reversing and taking the complement (e.g. http://arep.med.harvard.edu/labgc/adnan/projects/Utilities/revcomp. html or http://www.bioinformatics.nl/cgi-bin/emboss/revseq or write your own code) of the your coding strand DNA. Explain what happened and why. Save (and submit) the nucleotide sequence of the template strand that corresponds to these amino acids as a FASTA format file.

```
##    X.c5662.1.Reversed..Schizosaccharomyces.pombe.chromosome.I..complete.sequence
## 1                    GATCACGTACATCACCTTGTAAGAATTTATCTGCAATAGTCCTTCGGTATTGTACATTGT
## 2                    TCCAAGCATAGTAAACTAACGATATCAAGTTTGCCTTTCTAGCCCATGACCTACAGTCAG
## 3                    AAGTGTAAGCCATATCACTGTCGGCATGTTCAAACTTTGTCAAACCACAAAATAAACAAA
## 4                    GTCCTTGAAATCGAATACGTAGTTTACATTCTCGCAAGTTGTGGTCGGCCTTGCCACATT
## 5                    TATAACAAGTAGATAAGCGTACGGGGCATGCTTTCCCGGTATGAGCACGAATTTCTGTGT
## 6                    CTGGGTTACCAAGAGTGCAACTTAGACATTCATCTTTATACACTCGAAAGAATCGCTCGA
## 7                    GTTGTGCTGAAATGTCAGTTGAGTCTACCCATTGTTTTTTGACGGCTTGCACACGACTTT
## 8                    TATACTCGGCGAAATGAATGGTACGAGAATCAACATCACCATCACCATCACCATCACCAA
```

```
## 9                    TACCAATACCAATACCAACAATACCACTCATACCTCCAACACTACCATGAACACGATTCA
## 10                   TGTCATGGTTGACGCCCTGTTGTACATTGTTCAAATGTTCATCGTATGTGCGTTTATGAC
## 11                   CACGACGATGGCTGAGCGACATGTCTTCCAAACCATACTTTAGATTCGTATTAGCATTGC
## 12                   CAAAGCTTTTCTTCCATGTTGGCGATACAGACGAAGGGAACAAAGTGATTTCCGATAATT
## 13                   CCGACAAAGACACTGGCGTAGTACTAGTAGTGTTAGTGTTCATAGCACTCATACCACTGC
## 14                   TATTCCCTGGCTGTGGTGGGGGGGAAGACGAAAACGATGCATTATAACGCGAATGTGTTG
## 15                   CAATGGCTGGTTTCGGTGTTTCTGGCAATGTCGGTTTCACTCCATACATCGTAGACACAG
## 16                   TTGATTCTTCACCAAGTAACGAATCTGAGCATCTTGAGCAGTAAACACAGTTAGCAAACG
## 17                   ATGCACAACATACACATTCGCCATCCATTTCACTTGCGAGAAACGATCGAACACACATCG
## 18                   TATTATCATTAAGAAAGTTTTTCATCGAATCCTCTACGTAGCTCGACCATGTAGAATCAT
## 19                   ATTTCTCGTAAAACAATGCTGCAATCGCATACTTGCCATCTCTTCCAGCTCGACCTGTCT
## 20                   CCTGTACATAATCCATAGATGAAGCTGGTAATCCATAGTGTACTACTAAACGCACTCCCA
## 21                   TATAGTTGATACCGAGTCCGAATGCCTTGGTAGCGATCATGATTCGTGTCTTCCCATTTG
## 22                   CATTTCGAAACGCGTCAAAGTTCATTTGTCGTTCTTCGTCGCTTACATCTCCTGTGTAAA
## 23                   TGGTTACATGGGTGTGAGCGAACAAGTCCGACTGGTGAAGACGACGATGAATGTATTCGA
## 24                   CATCCTTTTTGGTTCGACAAAAGATGATAGCACGTCCATCACCTTCAAAGACCTTTGTTC
## 25                   GTTTCATCAACGTTCGCAAGTCCAACAAAAACTCGGTATTATCATAGGGATGCAAAAAGT
## 26                   AAAAGATGTTTTCCCGTGCAGTAGACGTTTCTCGTAAAACGTAAAAGTTTGTACAAAACG
## 27                   TTTGTCTGGCAACCATCTCTAGTTGCCTGGGAAAGGTGGCACTCAACAAGTGCAATGGTG
## 28                   CATACAAGCCAGACAACCGCGATGCTCTCGACAAAGCCGTTCTCCATGCTCCACTGGTGA
## 29                   GTAACAAATGTGCTTCATCAATTACCACTCGTGCCAAACGACCAAGTGTTGCCAAACTTT
## 30                   CAAAAAACCGAAGACCACTGTTGGTTAATGCTGACTCGTATGTCAAGATAAACAAATCGG
## 31                   GAAGTTGCGTCTCTAAAGTCAATCGTACATCTTTGAATGCGGTCCAATTCCCCGAACAAA
## 32                   CAAGCAGTCCCTTTTCATTCACTCTGAGCATCATATCTTGCCGTAACGACATCATGGGCA
## 33                   CTAGAACCAGCGTGACCATATTCATCACCTTTCCTGGGGTTTGTCTTTTTTTTTTCGATGA
## 34                   GCGCCGGTATCAAAAACGACAAAGACTTTCCACCTCCAGTTGGGAGTACGGTAATCAAAT
## 35                   TCATACGATTTAAAAGAGAAAAATAAACCGATTGAAATTGTTTCAGCGAACGGAATTTCG
## 36                   CTTCGAGACCATAATATTGAGACAATGCCCAACAACAATGTTGAACAGAAGTTTGAGATG
## 37                   GTTGAAGAAGTCCTGGGCGGTAGTGAAATGCGGTGGGCGCGGTGGGCGTAGTGGCCAATT
## 38                   GGTCTTGTAAGTCATGAGAGTCTTGATTTACAACGGGCTGTGTCATTTCAGCCTTTACCT
## 39                   CGTCCTTGTTCTTCTCCTCATTTGTCTTGTCCTTTTCCTTGTCCCCCTCCACCATTTGTT
## 40                   CACCTACCAACCTTGCCAGCTTCTCATTCAGCGCTTCCTCCAATTGATCAACTCGCTTGA
## 41                   TGAATGGGTGCTTTGTTTCCACCAACAGTCCATGGGTCGGGTTATCTCGAATCTGTAATA
## 42                   GTTCCTGCCACTTATAGCTTGCACGAAAAAAGTTGGCGAAATAATCCGACGGCAGATAAT
## 43                   GCAAGTTGTCCATAGTGCGCCCATAGATGTAGAGCGCTGTGTTTCGTGTATGTCCAGCCT
## 44                   GTAAATCGAAATATGATTCCCTCGTCATTTCTTCCTCGATGTTTTTTTCTTTAAAATAGT
## 45                   GAGCAATGTGCCTGTAGTTTCGAAATCCCAACGGTTTTTGAATGCATTGGTAGGTGGCCT
## 46                   TTGGAAAAATTCGATACGGTAAATCTCTTTGCAATCGTTCGCCAGCAATCACAAACATGA
## 47                   AATCCAAGTATACAGCTACTTTCGACCTATCAGCGGTTGTCACATACTTCATCAATGCTT
## 48                   CCAATGGTCGAACCAAAACATAGTACCGAAGTGCTAAAATGGAAAGCGGCTCAGAAAGAA
## 49                   ACCTGGGGATTGGCTTTTCAGCAAACTTCATATTACGAGTCTTATCGTATCTGCTGTAAA
## 50                   TCATCAGCCTTCCAAACATCAAATACAATTCGCGAGTCTTATACTTGCCATTCCGCAAAG
## 51                   TCCAATACACCATCTCTTGTGCTCTGGCTGGTTGTCCAGCAGATATGTACACTAGACAAA
## 52                   ACAATAACTCATTGAAGATAGATGCCTTTGTCTCATACTTTTTAATAGCAGATTTAGACC
## 53                   AATGTAATTTGAGAAACACATCATTGTCATTATCATTAGAGTTGTCGTTGTAGTGGTTGC
## 54                   TATAGTTGTCGCACTCATTTCTTTGCTGATCACTTACTCCACCAAACAACTTTGTCAGAT
## 55                   CCTTCTTAGACATCGATGGGAAAAACCGATGACAAAGTTCACTGTTCTTAAGTATTGTTC
## 56                   GTAACAGCACACTATGAGATGAACGTATCTCTATCGACTCTTTAAAGAATGAATATCCCA
## 57                   ATTTACTATTGTTTATGTCATCTCCAATTGACCTTAGTAACGAAATGGGTAATAGTTTTT
## 58                   CCACATTCATATCAAAACATAGCTCTTTTAACAAGTCCTGAACTTTGGCAAACATCCCCT
## 59                   CATACAGAGCTGTCAGCTCACTCAAGTCCAATCGCACCATGTTATTGAATGCATCTCCAA
## 60                   TGCACGGGTACATAGAATTACTTCGTGGCTGAAACAATACGGTACTTGGGCTTAGTCCTT
## 61                   GCAGTGTATTTGCATACCCTTCCGTTCGAATGATATCCCTAACGACTTTTTGGGTTTCGT
## 62                   AAAGAACATCATATAAATTTCTTTGACCAAGCGTGATAAAGCAAAACACATGAGCATAAG
```

```
## 63                  CAATAGAACCAGCGGTTTGAAACGATGGAATGGACAACAAGTCGATTTGTTTTTCATCGA
## 64                  TCAATCGTTGTAACTCACATGCTGCTACAAGACGACACGTGTATATTAGGGGTGCAAAGC
## 65                  AGGCAGAGATTTCATAGGCATATTTTGAAGTACCATCTTTATGCAAAGACATGCATGCAA
## 66                  GAAACTCCATGATTTCCAAGTCTTGAACATTTTTAGAAAAATCGTATTGTTGCAGAAAAG
## 67                  CCAATTTTAATGCCTCGTGCAGTTCCAAGTATGCCAGCTTATCATCATCGTCAGTATTGG
## 68                  TGTTGTTATTGTTATTGTTATTGTTACTGTCTCTGTCATTGTCATTGTCATTGTCATTGT
## 69                  CGTCTTCAGCACTATCAAAGTTGGCGACGTTGCTGATGTTACTGCTGCTGTTACTGCCAC
## 70                  CTTCTTTTCTCTCCACTGTTGCTCCACACATTGCTTCCAAGTGTCTTCTTATCCCCGAGT
## 71                  CCGCTGATGGACGACGTAACAGAAATATTAGGTATAGTGCTGCACTTTGAGCATATCTCA
## 72                  CAACTGTATCCTTTTGCTCTAAGCGACAAAGTCCTCGCTCTTGTGCACTATAGCTTCCTC
## 73                  CTTGAGTAAACATTCGTCTCATGCCCATGGTCATTCCACTTAGATGACGGAATCCTCGAT
## 74                  AGTAATAACGAATCAACTGGTTTACAATGGTATCCAAAGTTACAATTCCATTCGTATATT
## 75                  CGAATCGAACCAGCTGATCCCAATTTTCTCCAAGTTTTTTCACAATTGCCGTCCAATTGA
## 76                  ATTGTTGATTTAAAAGATTTTTCTCTCGTGCATTGACAATCTCTTCCTCTTCCTCTTCAT
## 77                  CGTTATCTTTTTCATCATCTTCAAGAATAGTGTTAGTGTTGTTGTCTGTTGCTTGGTCAA
## 78                  CATCATCTTGTCGTTTTTTATTTTCACCTTTGTCTTGCTGCTCCCGCCGCTGTTGATTCT
## 79                  TCTCCATAAATTTTGAAAAGCAAAAGTCCACCATTCTTTCCTTGATGTTTTGTGCATCTT
## 80                  CTTTCTCCTCCCCCTCCTCCTCTTCTTCTTCTCCTCCATTCAAAGGTACATAATCGA
## 81                  CTTTGAAGAACTGACATCGTTGAGCATACTTGTTCTTAACCGTTTGCAAAGCTGTTCGAC
## 82                  GTATACAATTTTCCAATTTTATAACTTTTCCATGAGTACGTCGAACATGATGGCGAAAGG
## 83                  TATCCATGATTGCATGTACGAATCCTGTTCCGTTTTTTGTGCACGGCACACACTCATATC
## 84                  CATTTAAAAGTACTGGTAGTCCTCTGATGTATGGGTATACATGGGTTTGAGAAGAATGAG
## 85                  TTTGTAAAACATTTGTTGGACTTTTTAACTTTAAAGTCCTTAGTTCTTGAAACCACAAAA
## 86                  GATCTTCGTTTAACTCTAGTTTATGCACAATTTGCATATGTTGCGCAGTGTGAATCACGT
## 87                  TTAACAAGCATTCACATTCTACGCACATCAAAGCATGAATAGACAAAATGGAGAGTCCAT
## 88                  AATTAGCCAACTTGTGATTCAAGTCAGTATAATTTGATGATGGGCTGTTGTCCAATTTGG
## 89                  TGTCATTTGTACTTTGCCAAATTAGCCGTGCAGTCCACAAGTCGTCATTTGCACTCGTAG
## 90                  TGAGAGTAGAGAGAGCAGAGAGAGCAGAGAGAGCGGGTAGTTGACGCTCCTTGGAAGAAT
## 91                  TGCAAGCCTCCTCACTGGTAGCCTCACTGGCAGCCTCAACGATGATTGGCTTATCCTTTT
## 92                  TCTCGTCCTGAACGAAAGAGCTGGGTTCATAGACGACGATATCATCGTCTTTACAAGTTT
## 93                  CGTCTCTGTATCGACATTCCTTGAAATGTTGAACGATCCTTTCTACATTATCAAATTGAG
## 94                  TTCCACATTGACAAGTAAAGCATCCGGCGATATCTCTAGCAGTTTTTGAAGCGACTTTAG
## 95                                                      CAATTTCTGAAGCGACGACCAT
```

In reverse compliment, The last base pair will be converted to first base pair and the direction will change from 5' to 3' to 3' to 5'. The above sequence is just a reverse compliment sequence.

2.5

Using the sequence shown in the record, give the nucleotide number range that corresponds to these amino acids (protein sequence). Find and report the stop codon in the nucleotide sequence. On which chromosome does the genomic sequence lie?
The nucleotide number range that corresponds to these amino acids using the sequence in record is MVVA and the number range is 5651 to 5662 however, its the first 12 (1 to 12) that would correcponds to MVVA in case we have selected "Show reverese compliment" on GenBank.

Question 3

3.1

Read up on C. elegans and in a few sentences describe it and why it is such an important organism for the scientific community.

C. elegans (Caenorhabditis elegans) is unsegmented, vermiform pseudocoelomate which lacks respiratory or circulatory system. The first research on it started in 1963 in australia. It is the only species which has connectome diagram. It does not require a sexual partner to reproduce. It is first multi cellular organism to have whole genome sequenced.

3.2

Use the nucleotide BLAST tool to construct a schematic diagram that shows the arrangement of introns and exons in the genomic sequence. In the BLAST tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch, choose database RefSeq Genome Database and remember that the species source of the genomic sequence is Caenorhabditis elegans. Use the 4 Genome Data Viewer button. Alternatively you may use https://wormbase.org/tools/blast_blat.

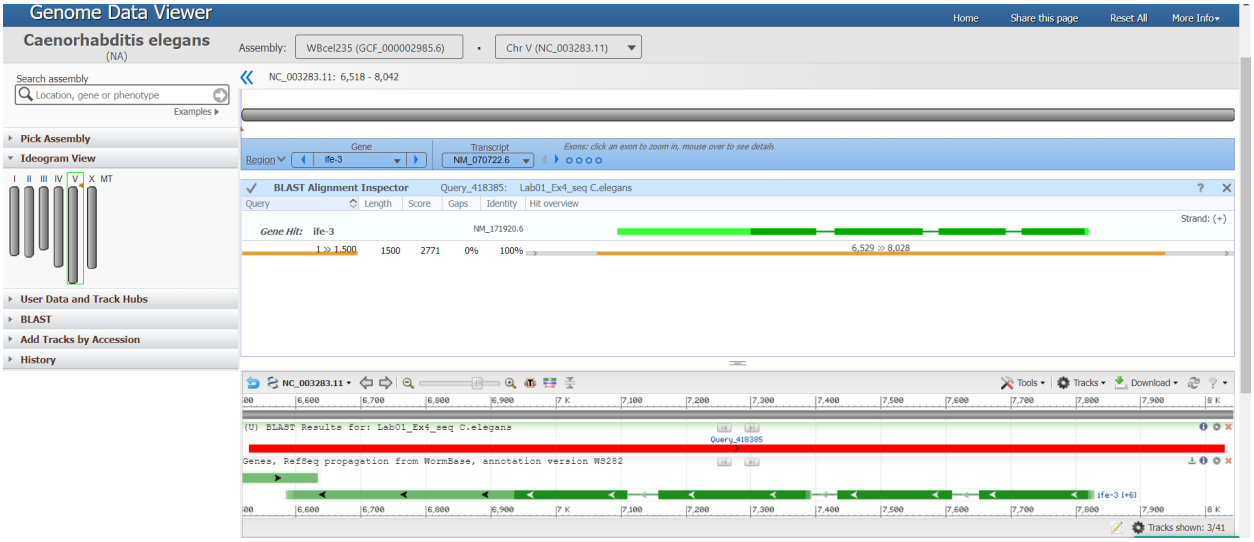

Figure 1: Schematic Diagram

3.3

Note the numbering of the sequences in the alignment (i.e. pairing of query and database sequences). Does the database genomic sequence progress in the same direction as the query sequence? What would happen if you reverse complement your query sequence (e.g. http: //arep.med.harvard.edu/labgc/adnan/projects/Utilities/revcomp.html or http://www. bioinformatics.nl/cgi-bin/emboss/revseq or write your own code) and do the search with such a reverse complemented sequence?

The database genomic sequence is shown below:

```
##                NC_003283.11.6529.8028.Caenorhabditis.elegans.chromosome.V
## 1  ATTTTTAAAAATGTACAAAATCAAACGCCCTACAAATCATGTGTGTGAAGAAGAATAATAACTAACATAT
## 2  CTATTTATATTTACCGAATAAATATATATTCATCAATTAACCTGAAGAACAAACGAATTCGGCTACAGGC
## 3  GTCGATCAGTCTCGAATCTAGTAACAACAAGAGAGCAATACGAAAACCGGTAAATCAATAGGGGGAAGCG
## 4  AAACAGTAGGTACAAATTGGAGGGGAAGCACCAATACATTAGGTGGGGGGTACGACTTGAAAAATGAGCT
## 5  GATTTTCGAATAGTTAAAGCGATGATCGTGTCCGAAAAACAGTTCATTTTTCAAGACAACATTGAGACTG
## 6  GGAGTACGGGGAAGCTCATTTACGGTGAGAGGAATTGGTGAGATCTTTAGAATATGCTTAAGGAGTTGGG
## 7  GTGGCTGGAGAAGTTCCTGTAGCCTCCGTGCCGGGATTCGATGGAGAAGTCGTTGCGGCTGGTCCCTTTT
## 8  CCTTCACTGGTGCTGGATCCTTGGCTGGAAGACATATGCGTGGCTTGACAGTCGATGAGGTGCGAGCCGA
```

```
## 9  CGAGTCCTTGTGAACTTCGTATCTGGAAATATTTTACTTAGATAGCAAATACTAAAATTGTAAAATTACC
## 10 TCAAAATCTCAGTATCCGGAATGCTCAATTTCTGCTTCAAAACCTGTCCGATGCGAAGATTGACATCATC
## 11 GCGAGTAGCATCACGAGTCCACAAGGAAACCTTGTCACCCTTTTGACGAACATTCACGACAGCTCCGCAG
## 12 ATGTAGTCTCCGTACTCGTCGAATTGCTCTCCAACAATAGCCATCAACAGCTCCAACCAGTAGTGATCGA
## 13 GCAATTGCGTTCTTCTCTGAAGCTTCTATGATTCATTGAATAAAATATATTTCTCAAAACGTACTTGCTT
## 14 ATCGACAACAACCAACCAACGTCCACCTTGAACGTTGTTGACGTCCTCCCACATTGGCTTGATTCCTTCC
## 15 TTGAACAAGTAATAATCGGATCCCCAGTTCAATCCTCCGGCAGACTGAATGTGATTGTACAGCGACCAGA
## 16 AGTCCTCGACAGTGTCGAAAAGTGAAACCATCTGGAAAAAATCGATAAAAGACGTATTTAAAAATCTTCT
## 17 ACCTTCAGACAATCCTCCCATTCCTTGTTACGGTCAGCTTTCAAGTACCAGAGAGCCCAGCGATTCTGGA
## 18 GGGGGTGTCTGGTGAGAAGCTCTGGAGGAACTGAAGCATCGGACGCATTCACATCGCCGGAAGCTGACAA
## 19 TGCTTTGTTTTCCGCTACGGATGTGCTCATTTAGCTGAAAATAGGTAATATTATATACGATTAGAGCTCG
## 20 GAAAACGATAAAATAGAGAAGAGTATGAATTTGGTTCAAATAACTCGGATTTTATAGGAAATTTTGTTTT
## 21 ACTGCACATTTTCGGCTAGTTTCCAAGCTTTTTAGATTTTTCAAGTGTAATTGGTAACATCGGGCACAAT
## 22                                          AAATTGATATTAAAGCTTGGAAAACAATAA
```

The reverse complemented sequence of the query sequence is:

```
##                             X.Lab01_Ex4_seq.Reversed..C.elegans
## 1  TTATTGTTTTCCAAGCTTTAATATCAATTTATTGTGCCCGATGTTACCAATTACACTTGA
## 2  AAAATCTAAAAAGCTTGGAAACTAGCCGAAAATGTGCAGTAAAACAAAATTTCCTATAAA
## 3  ATCCGAGTTATTTGAACCAAATTCATACTCTTCTCTATTTTATCGTTTTCCGAGCTCTAA
## 4  TCGTATATAATATTACCTATTTTCAGCTAAATGAGCACATCCGTAGCGGAAAACAAAGCA
## 5  TTGTCAGCTTCCGGCGATGTGAATGCGTCCGATGCTTCAGTTCCTCCAGAGCTTCTCACC
## 6  AGACACCCCCTCCAGAATCGCTGGGCTCTCTGGTACTTGAAAGCTGACCGTAACAAGGAA
## 7  TGGGAGGATTGTCTGAAGGTAGAAGATTTTTAAATACGTCTTTTATCGATTTTTTCCAGA
## 8  TGGTTTCACTTTTCGACACTGTCGAGGACTTCTGGTCGCTGTACAATCACATTCAGTCTG
## 9  CCGGAGGATTGAACTGGGGATCCGATTATTACTTGTTCAAGGAAGGAATCAAGCCAATGT
## 10 GGGAGGACGTCAACAACGTTCAAGGTGGACGTTGGTTGGTTGTTGTCGATAAGCAAGTAC
## 11 GTTTTGAGAAATATATTTTATTCAATGAATCATAGAAGCTTCAGAGAAGAACGCAATTGC
## 12 TCGATCACTACTGGTTGGAGCTGTTGATGGCTATTGTTGGAGAGCAATTCGACGAGTACG
## 13 GAGACTACATCTGCGGAGCTGTCGTGAATGTTCGTCAAAAGGGTGACAAGGTTTCCTTGT
## 14 GGACTCGTGATGCTACTCGCGATGATGTCAATCTTCGCATCGGACAGGTTTTGAAGCAGA
## 15 AATTGAGCATTCCGGATACTGAGATTTTGAGGTAATTTTACAATTTTAGTATTTGCTATC
## 16 TAAGTAAAATATTTCCAGATACGAAGTTCACAAGGACTCGTCGGCTCGCACCTCATCGAC
## 17 TGTCAAGCCACGCATATGTCTTCCAGCCAAGGATCCAGCACCAGTGAAGGAAAAGGGACC
## 18 AGCCGCAACGACTTCTCCATCGAATCCCGGCACGGAGGCTACAGGAACTTCTCCAGCCAC
## 19 CCCAACTCCTTAAGCATATTCTAAAGATCTCACCAATTCCTCTCACCGTAAATGAGCTTC
## 20 CCCGTACTCCCAGTCTCAATGTTGTCTTGAAAAATGAACTGTTTTTCGGACACGATCATC
## 21 GCTTTAACTATTCGAAAATCAGCTCATTTTTCAAGTCGTACCCCCCACCTAATGTATTGG
## 22 TGCTTCCCCTCCAATTTGTACCTACTGTTTCGCTTCCCCCTATTGATTTACCGGTTTTCG
## 23 TATTGCTCTCTTGTTGTTACTAGATTCGAGACTGATCGACGCCTGTAGCCGAATTCGTTT
## 24 GTTCTTCAGGTTAATTGATGAATATATATTTATTCGGTAAATATAAATAGATATGTTAGT
## 25 TATTATTCTTCTTCACACACATGATTTGTAGGGCGTTTGATTTTGTACATTTTTAAAAAT
```

The database genomic sequence moves in the same direction. The direction is from 1 to 60, 61 to 120, etc and 6529 to 6588, 6589 to 6648.

3.4

On what chromosome and what position is the query sequence found?

The position in the query sequence is 6529 to 8028.

3.5

Extract the DNA code of each exon and using transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/) find the protein code of the gene. You can also use blastx (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome or https://wormbase.org/tools/blast_blat) to obtain protein sequences. How do they compare to your translation?
Tbhe protein code of the gene using blastx is:

```
## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
## incomplete final line found by readTableHeader on 'out2.txt'
```

```
##                                                              Protein.Sequence
## 1 NKEWEDCLKMVSLFDTVEDFWSLYNHIQSAGGLNWGSDYYLFKEGIKPMWEDVNNVQGGRWLVVVDKQKL
## 2 QRRTQLLDHYWLELLMAIVGEQFDEYGDYICGAVVNVRQKGDKVSLWTRDATRDDVNLRIGQVLKQKLSI
## 3                                                                   PDTEILR
```

DNA code of each exon using transeq:

```
##                                                      Reverse
## 1 LLFSKL*YQFIVPDVTNYT*KI*KAWKLAENVQ*NKISYKIRVI*TKFILFSILSFSEL*
## 2 SYIILPIFS*MSTSVAENKALSASGDVNASDASVPPELLTRHPLQNRWALWYLKADRNKE
## 3 WEDCLKVEDF*IRLLSIFSRWFHFSTLSRTSGRCTITFSLPED*TGDPIITCSRKESSQC
## 4 GRTSTTFKVDVGWLLSISKYVLRNIFYSMNHRSFREERNCSITTGWSC*WLLLESNSTST
## 5 ETTSAELS*MFVKRVTRFPCGLVMLLAMMSIFASDRF*SRN*AFRILRF*GNFTILVFAI
## 6 *VKYFQIRSSQGLVGSHLIDCQATHMSSSQGSSTSEGKGTSRNDFSIESRHGGYRNFSSH
## 7 PNSLSIF*RSHQFLSP*MSFPVLPVSMLS*KMNCFSDTIIALTIRKSAHFSSRTPHLMYW
## 8 CFPSNLYLLFRFPLLIYRFSYCSLVVTRFETDRRL*PNSFVLQVN**IYIYSVNINRYVS
## 9                                        YYSSSHT*FVGRLILYIFKN
```

3.6

Hovering over an exon you should see links to View GeneID and View WormBase. These point to pages with more information on the gene. Follow them and write a few sentences about the gene.
The gene symbol is ife-3. The gene belongs to eukaryotic translation initiation factor 4E family (eIF-4E) with eukaryotic cellular mRNA. In humans it is encoded as EIF4E3 gene.The exon count is 4.

Reference

1. https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle

2. https://www.sciencedirect.com/topics/neuroscience/hardy-weinberg-principle
3. https://biology.andover.edu/Joomla/index.php/andover-biology-department-textbooks/biol-58x-sequence-advanced-biology-textbook/evolution/436-the-hardy-weinberg-equilibrium

4. https://iupac.qmul.ac.uk/AminoAcid/A2021.html
5. https://en.wikipedia.org/wiki/Caenorhabditis_elegans
6. https://en.wikipedia.org/wiki/EIF4E3

Appendix

```
knitr::opts_chunk$set(echo = TRUE)
#Given that
n<-1000   # population
mm<-357   # number of mm
mn<-485   # number of mn
nn<-158   # number of nn
P_mm<-mm/n  #proportion of mm in population
P_mn<-mn/n  #proportion of mm in population
P_nn<-nn/n  #proportion of mm in population

#Now calculting p, q, p2 and q2 (i.e all parameters) for Hardy-Weinberg equilibrium

#proportion of N and M in the alleles are represented by p and q respectively :
# one way to calculate
p<-P_mm+(0.5*P_mn)
q<-P_nn+(0.5*P_mn)

#another way to calculate p and q
#total no of alleles is 2n
total<-2*n
p<-(mm*2+mn)/total
q<-(nn*2+mn)/total

#cross check
p+q  # p+q =1

#Now, calculating p_square, q_square and two_pq

p_square<-p^2
q_square<-q^2
two_pq<-2*p*q



#chi-square goodness of fit test (one method)

chi_test = chisq.test(c(P_mm, P_mn, P_nn), p=c(p_square, two_pq, q_square))
print(chi_test)



sequ <- read.delim("out.txt")
print(sequ)
sequ <- read.delim("outseq.txt")
print(sequ)
sequ <- read.delim("out1.txt")
print(sequ)
sequ <- read.delim("outseq1.txt")
print(sequ)
sequ <- read.delim("out2.txt")
print(sequ)
```

```
sequ <- read.delim("outseq2.txt")
print(sequ)
```