

732A51 Bioinformatics Lab 1

Raymond Sseguya, Martin Smelik, Duc Duong

2018 M11 7

Task 1

Task 1.1

diploid population = Aa,Aa = 2N (parents)

After random mating for first generation: children = AA, Aa, aA, aa

Total new first generation population = 4

Number of AA homozygotes = 1

Number of aa homozygotes = 1

Number of Aa or aA heterozygotes = 2

Proportion of AA homozygotes = $1/4 = 0.25$

Proportion of aa homozygotes = $1/4 = 0.25$

Proportion of Aa or aA heterozygotes = $2/4 = 0.5$

The proportion AA Homozygotes : Aa Heterozygotes : aa homozygotes is

0.25: 0.5 : 0.25

Again after random mating for first generation:

Probability of getting A allele = $3/6 = 0.5 = p$

Probability of getting a allele = $3/6 = 0.5 = q$

The proportion AA Homozygotes : Aa Heterozygotes : aa homozygotes is also

square of p : 2 times p times q : square of q

$(0.5)^2 : 2 \text{ times } 0.5 \text{ times } 0.5 : (0.5)^2$

0.25: 0.5 : 0.25

This satisfies the Hardy Weinberg equilibrium.

Task 1.1.b

The probability of getting A allele and a allele will remain the same with continued random mating therefore the Hardy Weinberg equilibrium will always hold.

Task 1.2

Total number of people = $357 + 485 + 158 = 1000$

Total allele population = $2 \text{ times } 1000 = 2000$

Total number of M is $2 \text{ times } 357 \text{ added to } 485 = 1199$

Total number of N is 2 times 158 added to 485 = 801
 Probability of getting M is 1199 out of 2000 = 0.5995 (assuming diploid)
 Probability of getting N is 801 out of 2000 = 0.4005 (assuming diploid)
 Creating vector of number of homozygotes and heterozygotes, R
 Creating vector of Probabilities of M and N alleles, S

```
R <- c(357, 485, 158)
S <- c(0.5995*0.5995, 2*0.5995*0.4005, 0.4005*0.4005)

chisq.test(R, p=S)
```

```
##
## Chi-squared test for given probabilities
##
## data: R
## X-squared = 0.099938, df = 2, p-value = 0.9513
```

The null hypothesis under the chi-square test for goodness of fit, that the population follows the Hardy Weinberg equilibrium, IS ACCEPTED.

Task 3

3.1

According to Wikipedia, *C. elegans* is being extensively used as a model organism. It was the first multicellular organism to have its whole genome sequenced, and as of 2012, is the only organism to have its connectome (neuronal “wiring diagram”) completed. The *C. elegans* genome contains an estimated 20,470 protein-coding genes.[95] About 35% of *C. elegans* genes have human homologs. Remarkably, human genes have been shown repeatedly to replace their *C. elegans* homologs when introduced into *C. elegans*. Conversely, many *C. elegans* genes can function similarly to mammalian genes.

3.2

3.3

```
Lab01_Ex4_seq   Reversed:   C.elegans   TTATTGTTTTCCAAGCTTTAATATCAATT-
TATTGTGCCCCGATGTTACCAATTACACTTGA      AAAATCTAAAAAGCTTGGAAAC-
TAGCCGAAAATGTGCAGTAAAACAAAATTTCTCTATAAA  ATCCGAGTTATTTGAAC-
CAAATTCATACTCTTCTCTATTTTATCGTTTTTCCGAGCTCTAA  TCGTATATAATAT-
TACCTATTTTCAGCTAAATGAGCACATCCGTAGCGGAAAACAAAGCA TTGTCAGCTTC-
CGGCGATGTGAATGCGTCCGATGCTTCAGTTCCTCCAGAGCTTCTCACC  AGACAC-
CCCCTCCAGAATCGCTGGGCTCTCTGGTACTTGAAAGCTGACCGTAACAAGGAA
TGGGAGGATTGTCTGAAGGTAGAAGATTTTTTAAATACGTCTTTTATCGATTTTTTCCAGA
TGGTTTCACTTTTTCGACACTGTTCGAGGACTTCTGGTTCGCTGTACAATCACATTCACTCTG
CCGGAGGATTGAACTGGGGATCCGATTATTACTTGTTCAAGGAAGGAATCAAGC-
CAATGT      GGGAGGACGTCAACAACGTTCAAGGTGGACGTTGGTTGGTTGTTGTC-
GATAAGCAAGTAC  GTTTTGAGAAATATATTTTATTCAATGAATCATAGAAGCTTCA-
GAGAAGAACGCAATTGC  TCGATCACTACTGGTTGGAGCTGTTGATGGCTATTGTTG-
```

GAGAGCAATTCGACGAGTACG GAGACTACATCTGCGGAGCTGTCGTGAATGTTTCGT-
 CAAAAGGGTGACAAGGTTTCCTTGT GGACTCGTGATGCTACTCGCGATGATGT-
 CAATCTTCGCATCGGACAGGTTTTGAAGCAGA AATTGAGCATTCCGGATACTGA-
 GATTTTGAGGTAATTTTACAATTTTAGTATTTGCTATC TAAGTAAAATATTTCCA-
 GATACGAAGTTCACAAGGACTCGTCGGCTCGCACCTCATCGAC TGTCAAGCCACG-
 CATATGTCTTCCAGCCAAGGATCCAGCACCAGTGAAGGAAAAGGGACC AGCCG-
 CAACGACTTCTCCATCGAATCCCGGCACGGAGGCTACAGGAACTTCTCCAGCCAC
 CCCAACTCCTTAAGCATATTCTAAAGATCTCACCAATTCCTCTCACCGTAAAT-
 GAGCTTC CCCGTACTCCCAGTCTCAATGTTGTCTTGAAAAATGAACTGTTTTTCG-
 GACACGATCATC GCTTTAACTATTCGAAAATCAGCTCATTTTTCAAGTCGTACCC-
 CCCACCTAATGTATTGG TGCTTCCCTCCAATTTGTACCTACTGTTTCGCTTCCC-
 CCTATTGATTTACCGGTTTTCG TATTGCTCTCTTGTTGTTACTAGATTTCGAGACT-
 GATCGACGCCTGTAGCCGAATTCGTTT GTTCTTCAGGTTAATTGATGAATATATATT-
 TATTCGGTAAATATAAATAGATATGTAGT TATTATTCTTCTTCACACACATGATTTG-
 TAGGGCGTTTGATTTTGTACATTTTTTAAAAAT

3.4

The query sequence is found on the 11th position

3.5