732A51 Bioinformatics Lab 1

Raymond Sseguya, Martin Smelik, Duc Duong

2018 M11 7

Task 1

1.1

Aa + Aa -> AA, Aa, Aa, aa p = 1/2 q = 1/2 pˆ2 = 1/4 = frequency (AA) qˆ2 = 1/4 = frequency (aa) 2pq = 1/2 = frequency (Aa)

The proportion will always be p = 1/2 and q = 1/2 and the population should not deviate from the equilibrium.

1.2

For inspiration I used wikipedia https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle We should first calculate the frequencies MM = 357, MN = 485, NN = 158

p = (2*357 + 485) / 2000 = 0.5995 q = 1-p = 0.4005

count(AA) = pˆ2n = 0.5995ˆ2 * 1000 = 359,4 count(Aa) = 2pqn = 480,1 count(aa) = qˆ2n = 160,4

chiˆ2 = (357-359,4)ˆ2/359,4 + (485 - 480,1)ˆ2/480,1 + (158-160,4)ˆ2/160,4 = 0.016 + 0.05 + 0.036 = 0.102

We know that the 5% significance level is 3,84 and that we can not reject the hypothesis that the population is in Hardy-Weinberg proportions

Task 2

2.1

A - Adenine T - Thymine G - Guanine C - Cytosin

2.2

M - Methionine V - Valine V - Valine A - Alanine

2.3

We can see this nucleotide sequence in gene.txt file.

2.4

The obtained sequence is different to the one we were provided but after reversing and taking the complement we got exactly the same sequence as we were provided, shown in outseq.txt file. This happend because every DNA has two components, following the rule that T, A and C, G are complementary. It also depends from which end the sequence start and that is why reversing is sometimes necessary.

2.5

number range: 1..5662 stop codon: ???? chromosome="I"

Task 3

3.1

C. elegans (Caenorhabditis elegans) is an organism of length about 1mm, mostly hermaphrodite which has some of the organ systems symiliar to large animals, although they don't have neither respiratory nor circulatory system. C. elegans is important for bioinformatics because it is very simple but at the same time it also have a nervous system. There has been many important research on this organism and according to the wikipedia, it was first multicellular organism with the whole genome sequenced and is the only organism to have its connectome completed.

3.2