7.03 Exam 3 Practice Problem

Problem 1
There is a rare recessive X-linked trait that affects 1/8000 males in a rabbit population. The mutation causes black spots on the white rabbits (normal phenotype is plain white).

(a) What is the value of q?

(b) In what proportion of matings would this trait affect half of the male and female offspring?

(c) Now assume only 20% of the affected rabbits survive. What is the value of S?

(d) By how much would the allele frequency change between the current generation of rabbits and the next generation?

(e) Now new mutations are introduced to the population, with a mutation rate of $5x10^{-3}$. What would the allele frequency q equal after a new steady state had been reached?

Problem 2
For a specific gene, there are only two possible alleles, A and a. Professor Regev makes a new population where f(A) = f(a) = 0.5. Assume no new mutations, heterozygous advantage or selective disadvantage.

a) According to the conditions stated in the problem will this population most likely be in Hardy-Weinberg Equilibrium or in a balanced polymorphism? What are the frequencies of the possible genotypes?

b) Let's say an individual with the genotype AA or Aa is normal, but someone who is aa unfortunately dies shortly after birth. What is s, the selective disadvantage? After an infinite number of generations, what will be the allele frequencies for A and a? Explain.

c) There has been a new vaccine that lets 40% of the individuals who are aa survive. What is the new s? Will f(A) and f(a) be the same as in part b after many generations? Explain. If they do change, what are the new frequencies?

d) Now assume that there can be mutations, where 'A' is mutated to 'a' at a rate of $\mu = 10^{-5}$. Assuming that we start from the beginning population where f(A) = f(a) = 0.5 and the vaccine is there, what will be the equilibrium allele frequencies of p and q?

Problem 3

You have two sequences that you want align

A. Using the Needleman-Wunsch algorithm to perform a global alignment of the following two nucleotide sequences:

GAATTC
GGATCGA

And using the following scoring scheme:
Match = +2
Mismatch = -1
Gap = -2 per position

You get the following Dynamic Programming (DP) matrix with an optimal path marked in red.

		G	G	A	T	C	G	A
	0	-2	-4	-6	-8	-10	-12	-14
G	-2	2	0	-2	-4	-6	-8	-10
A	-4	0	1	2	0	-2	-4	-6
A	-6	-2	-1	3	1	-1	-3	-2
T	-8	-4	-3	1	5	3	1	-1
T	-10	-6	-5	-1	3	4	2	0
C	-12	-8	-7	-3	1	5	3	1

What is the alignment indicated by this matrix?

What is the score for this alignment?

B. Using the Smith-Waterman algorithm to perform a local alignment of the above nucleotide sequences using the same scoring scheme you get a DP matrix with an optimal path marked in red that looks like this

		G	G	A	T	C	G	A
	0	0	0	0	0	0	0	0
G	0	2	2	0	0	0	2	0
A	0	0	1	4	2	0	0	4
A	0	0	0	3	3	1	0	2
T	0	0	0	1	5	3	1	0
T	0	0	0	0	3	4	2	0
C	0	0	0	0	1	5	3	1

What is the alignment indicated by this matrix?

What is the alignment score?

What is the score of the following alignment of these sequences using the above scoring rules?

```
GAATTC---
GGAT--CGA
```

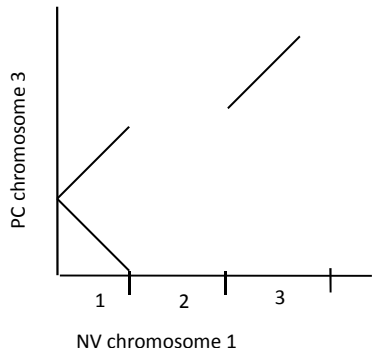
C. What is the difference between the local and global alignments in this example?

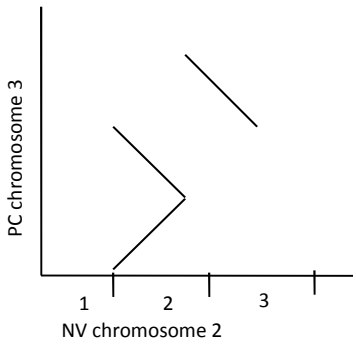
D. You find two proteins that you suspect have a common protein domain because they both can bind to the same DNA sequence. You wish to determine what this protein domain is that allows them to do this. Having honed your alignment skills as a 7.03 student you decide to align the two sequences and look for regions of high similarity. What kind of alignment (global or local) would you perform? Explain.

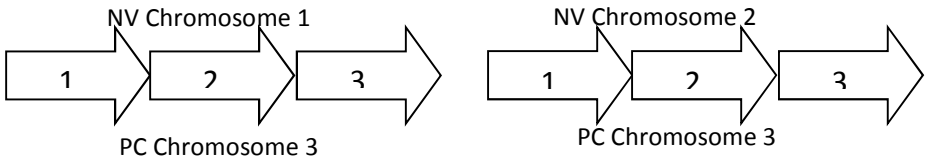
Problem 5

You are attempting to identify coding regions in a newly sequence genome of a sea anemone species *Protantheae carlgren* (PC). At your disposal is relatively well annotated genome of a model organism, the starlet sea anemone, *Nematostella vectensis* (NV).

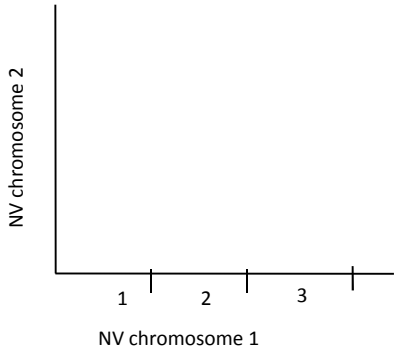
a. You begin with a large-scale chromosomal comparison between the two species. Underneath each dot plot, draw in the PC chromosome 3 arrangement relative each NV chromosome given. Use arrows to indicate sequence direction and number the segments accordingly (1, 2, 3).







b. Based on the information in the graphs in part a, fill in the following dot plot comparing NV chromosomes 1 and 2.



c. You discover three sequence segments in PC with similarity to **start** of a known protein coding region in NV (indicated by red text). Based on the alignments shown, choose the PC sequence that most likely represents a truly homologous, viable coding region and state your reasoning. Note that the non-template DNA strand is given for each sequence so consider only forward reading frames, and it is **not** necessary to translate any sequence.

Unaligned NVsequence:

```
5' CCATGACCTTCGACTCAGTCATCACTCTTGATGAT
    (start)
```

PC sequence 1 alignment:

```
NV: 5'CCATGACCTTC__GACTCAGTCATCACTCTTGATGAT
PC: 5'CCATGACCTTCGGGACTGA_TCATCACTCTGCTTGAT
      ***********  **** * **********   ****
```

PC sequence 2 alignment:

```
NV:  5'CCATGACCTTCGACTCAGTCATCACTCTTGATGAT
PC:  5'C_ATGACGTT_TAC__AGTAATAACCCTAGACGAT
       * ***** **  **  ****** ** ** ** ***
```

PC sequence 3 alignment:

```
NV:   5’CCATGACC_TTCGACTCAGTCATCACTCTTGATGAT
PC:   5’CCATGAGCTTTC_ACGCAGACTTGTCTCTTCATGTT
        ******** *** ** *** ***  ***** *****
```