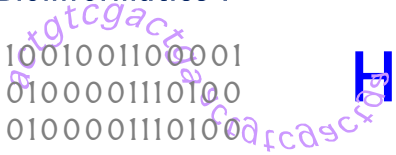

Homework Five

Please hand in the solutions to the following problems on Tuesday, November 25, 2014. Hand in a hard copy (required) and a USB (optional) or CD (optional) containing your solutions.

Problem One Problems from the textbook – Chapter Seven

1) When analyzing sequences with evolution in mind, is it the differences between them that we need to quantify and score?
2) What is the purpose of the phylogenetic tree representation?
3) a) What are species trees?
   b) How are they constructed?
4) Is the evolutionary history of a set of related genes always the same as that of the species from which the genes were selected?
5) a) What is a speciation event?
   b) How is it represented in a species trees?
6) What does the root represent in rooted trees?
7) What is the major task of phylogenetic tree reconstruction?
8) What is difference the between the following types of phylogenetic trees:
   a) cladograms
   b) additive trees
   c) ultrametric trees
9) a) What is meant by bootstrap analysis?
   b) How can this analysis be used to construct condensed trees?
10) What are the two conditions that would have made phylogenetic tree reconstruction from a set of homologous sequences considerably easier had they held during sequence evolution?
11) Where do most mutations that are retained in DNA come from?
12) What is the difference between synonymous and nonsynonymous mutations?
13) When is it useful to remove the third codon sites from the data before any further analysis?
14) What is the key assumption that is made when constructing a phylogenetic tree from a set of sequences?
15) Explain the process of gene loss. Does gene loss occur solely because of gene duplication?
16) What is meant by homoplasy?

17) What is meant by "horizontal gene transfer" (also known as lateral gene transfer")? Why is it called "horizontal"?
18) What are syntenic regions? Are they easily detected? Explain.
19) When comparing sequences from two closely related species, which regions will convey useful information for the construction of phylogenetic trees?
20) The analysis of which genomic sequence led to the discovery that prokaryotes comprised two quite distinct domains?

Problem Two

a) "Multiple founder effects and geographical clustering of BRCA1 and BRCA2 families in Finland" by Laura Sarantaus et al. appeared in the "European Journal of Human Genetics" in 2000.

The upper half of Table 1 of the article gives the mutations of the BRCA1 gene reported in Finland and also states whether the each mutation is confined to Finland or has been reported elsewhere in the world.

Table 1 Mutations and phenotypes of the *BRCA1* and *BRCA2* families in Finland

Gene and mutation	*No. of families in Finland*	*No. of female breast cancer cases[b] (mean age at dg)[c]*	*No. of ovarian cancer cases[b] (mean age at dg)[c]*	*Reported outside Finland*
BRCA1				
Ex 11, 1924delA	1	1 (44)	–	No
Ex 11, 2803delAA	2	3 (56)	2 (62)	the Netherlands[30]
Ex 11, 3604delA[a]	6	7 (45)	10 (46)	Belgium, the Netherlands[30]
Ex 11, 3744delT[a]	8	7 (45)	10 (49)	Sweden[10]
Ex 11, 3904C→A	1	3 (49)	3 (59)	Yes[3]
Ex 11, 4153delA	1	1 (32)	1 (48)	Latvia, Poland, Russia, Sweden[3,28,37]
Int 11, 4216nt-2A→G[a]	9	24 (43)	8 (52)	No
Ex 13, 4446C→T[a]	3	23 (46)	8 (53)	Belgium, Canada, France, UK, USA[3,28,29]
Ex 17, 5145del11	1	4 (37)	–	No
Ex 20, 5370C→T[a]	3	12 (49)	3 (67)	Austria[31]
Ex 20, 5382insC	1	2 (57)	1 (40)	Austria, Belgium, Canada, France, Germany, Hungary, Israel, Italy, Latvia, the Netherlands, Russia, UK, USA[3,28,37]

For each of the following two mutations, locate the mutation on the sequence and explain what consequences on the protein it might have. Clearly give details as was done with Problem 4 of HW3.

1) Ex 13, 4446 C → T
2) Ex 20, 5382 ins C

b) "Novel Germline BRCA1 and BRCA2 Mutations in Breast and Breast/Ovarian Cancer Families from the Czech Republic" by Eva Machackova et al. appeared in "Human Mutations" in 2001.

The article mentions four frame shift mutations in BRCA1 and BRCA2. We are going to study the two mutations reported in BRCA1.

For each of the following two frame shift mutations, locate the mutation on the sequence and explain the consequences on the protein. Clearly give details as was done in class for similar problems.

1) 3761-3762 del GA
2) 2616-2617 ins AAGTATCCAT

Problem Three

"Characterization of BRCA1 and BRCA2 splicing variants: a collaborative report by ENIGMA consortium members" by Mads Thomassen et al. appeared in the "Breast Cancer Research Treatment" in 2011.
Read the article and answer the following questions:

1) What is ENIGMA?
2) What is their goal?
3) What does this study demonstrate according to the end of the abstract?
4) What is BIC?
5) Are bioinformatics tools, such as algorithms implemented in web-based programs that predict splicing effects of nucleotide variants, sufficient for predicting splicing? Why?
6) What is the Splicing Working Group?
7) Name the four bioinformatics prediction algorithms used in this work.
8) What is HGVS?
9) What are cryptic sites?
10) The splice site predictor: NNSplice, that can be found at: http://www.fruitfly.org/seq_tools/splice.html is one of the bioinformatics tools that was used in the findings reported in the article.
    - Go to NCBI and obtain the MOG gene (accession number Z48051).

    a) Use NNSplice with the MOG gene to create a table for the "Donor site predictions" given by the program and explain, under Comment/Validation, whether you think that the splice site is authentic or not. Explain why you chose Yes or No. Note: your table might have fewer/more rows.
    b) Hint: Use different values (not the default of 0.4) for "score for 5' splice site"

Number	Start	End	Score	Comments/Validation
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

c) Same as a) but for the "Acceptor site predictions"

Number	Start	End	Score	Comments/Validation
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				