Spring 2014 Genomics Exam #1
Genomic Sequences

There is no time limit on this test, though I don't want you to spend too much time on it. I have tried to design an exam that will take less time that exams in the past. You do not need to read any additional papers other than the ones I send to you. There are 4 pages, including this cover sheet, for this test. There are no Discovery Questions on this exam. You are not allowed discuss the test with anyone until all exams are turned in at 10:30 am on Wednesday February 12. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 10:30 am ON WEDNESDAY FEBRUARY 12**. You may use a calculator, a ruler, your notes, the book, and the internet. You may work on this exam in as many blocks of time as you want. Submit your electronic version before 10:30 am (eastern time zone).

The **answers to the questions must be typed in a Word file and emailed to me as an attachment**. Be sure to backup your test answers just in case (I suggest a thumb drive or other removable medium). You will need to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain. Remember to explain your thoughts in *your* own words and use screen shots to support your answers. Screen shots without ***your*** words are worth very few points. Support your answers with data using screen shots liberally.

DO NOT READ or DOWNLOAD ANY PAPERS FOR THIS EXAM. You may search and read abstracts. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS. Spell out your logic for each answer.

-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please type):

Write out the full pledge and sign (electronic signature is ideal):

How long did this exam take you to complete?

20 pts

1) I want you to analyze some sequences. All of them can be found in the Word file called "Mystery_One.docx".

a) Translate **>First_Sequence**. How many amino acids does the encoded protein have?

b) From what species was this sequence taken? Support your answer with evidence.

c) I happen to know that **>Second_Sequence** evolved from **>First_Sequence**. Analyze **>Second_Sequence**. List and describe all the mutations you detect. Support each mutation you describe with the evidence you used. Your analysis should provide the most parsimonious answer.

d) Search NCBI to find a tabular presentation of some of the known SNPs in the original gene for First Sequence. Be sure you are looking at the correct species (see part b above). Your answer must include a screen show showing multiple SNPs and the effects they would have on the encoded protein.

20 pts

2) Over the last few years, we have sequenced more and more Neaderthal DNA. These questions pertain to those findings.

a) **Neander Fig1a** shows the analysis of DNA from 6 different fossil bone samples. Which fossil would you use for additional studies and why? Support your answer with data.

b) From an abstract: "Neanderthals are the extinct hominid group most closely related to contemporary humans, so their genome offers a unique opportunity to identify genetic changes specific to anatomically fully modern humans. Direct high-throughput sequencing of a DNA extract from this fossil has thus far yielded over one million base pairs of hominoid nuclear DNA sequences. Comparison with the human and chimpanzee genomes reveals that modern human and Neanderthal DNA sequences diverged on average about 500,000 years ago." What can you conclude from **Neander Fig1b**?

c) From the figure legend of **Neander Fig1c:** "a, Individual maps; the marginal probability of Neanderthal ancestry for one European-American, one east-Asian and one sub-Saharan-African phased genome across chromosome 9. b, Population maps; estimated the proportion of Neanderthal ancestry in European individuals (red) and east-Asian individuals (green), averaged across all individuals from each population in non-overlapping 100-kb windows on chromosome 9. The black bar denotes the coordinates of the centromere. The plot is limited to segments of the chromosome that pass quality filters. CEU, residents of Utah, US, with northern and western European ancestry; CHB, Han Chinese in Beijing, China; LWK, African Luhya in Webuye, Kenya." Interpret panels a and b from this figure. Support your answer with data from the figure.

d) Consider **Neander Fig1d** and summarize the results. Support your answer with data from the figure.

15 pts

3) This question focuses on the biggest news in biology for several years. I suspect this will win a Nobel prize not long from now.

a) Look up "STAP cells" and briefly summarize what they are. Provide me with a link to your source(s).

b) What genomic evidence did the investigators use to convince skeptics that they had converted mature, differentiated cells into stem cells? Provide citation for your source(s).

c) What sort of changes would you expect to see at the genome level if you compared the original cells to the STAP cells? Do not talk about protein or mRNA differences – focus on the genome.

25 pts

4) The first two snake genomes (Burmese Python and King Cobra) were sequenced in December, 2013. Answer the following questions.

a) Look at **Snake Fig1** (MKO = mouse knock out phenotype). Summarize what you see about snakes in general as well as each species of snake.

b) **Snake Fig2** focuses on snake venom that evolved in king cobras. (C) … the three-finger toxin gene family (D) other pathogenic toxin families of venom-expressed genes, and (E) ancillary toxin families after the split of the Burmese python from the advanced snakes. Compare and contrast evolution of the three toxin gene families.

c) Where are Bermese Pythons found in nature outside their normal range? Are they having an impact on their new location? Support your answer with data including a peer-reviewed journal citation.

d) What would you do next to determine if the Burmese Python is evolving in its new habitat? Design the experiment you would like to perform if expense and human resources were not limiting.

e) Find the Homo sapiens kallikrein 1 gene in a human genome browser. What compounds bind to human kallikrein 1? Support your answer with screen shots.

f) What is kallikrein 1's normal function and what human tissues transcribe this mRNA? Support your answer with screen shots. How could this function be related to snake venom?

g) How could the cobra's kallikrein 1 gene evolve to be venomous in king cobra if snakes use kallikrein 1 the same way humans do?

20 pts

5) The final question is a series of mildly related parts.

a) Look at **Seq Fig5a** and evaluate the diagram. This is a hypothetical case of one chromosome with three distinct regions (I, II and III). The short segments under the long contiguous piece of DNA are sequencing reads from a 454 machine. On top is the assembly from algorithm #1 and below is the assembly from algorithm #2. Summarize what algorithm #2 did by mistake. Explain how this mistake could take place.

b) Some authors stated, "hemizygosity of a 600-kilobase (kb) region on the short arm of chromosome 16 causes a highly penetrant form of obesity that is often associated with hyperphagia and intellectual disabilities. The corresponding reciprocal duplication is associated with being underweight. The reciprocal impact of these 16p11.2 copy-number variants indicates that severe obesity and being underweight could have mirror aetiologies, possibly through contrasting effects on energy balance." Do your best to determine how many distinct genes are in this region of the human genome. Explain to me how you determined your answer. Support your answer with screen shots.

c) Look at **Seq Fig5b** which shows GWAS data plotted on negative $\log_{10}$ p-value on the Y-axes. They were looking for genes that affect systolic blood pressure (SBP) and diastolic blood pressure blood pressure (DBP). The horizontal dotted line is $p = 2.5 \times 10^{-8}$. Pick one gene to study and tell me why you would choose it. What is the function of the encoded protein for the gene you chose (use GO terms)?

d) Look at **Seq Fig5c**. What can you conclude about the microbiome on humans and their pet dogs? The left panel shows the average unweighted UniFrac distance between adult dog-owners and their dogs (blue), between dog-owners and other (not their own) dogs (red), and between adults who do not own dogs and dogs (green). The right panel shows the number of phylotypes shared for the same categories. Comparisons are labeled on the y-axis such that the first body site listed corresponds to the dog and the second site corresponds to the human. Mean ± 95% CI shown. The presence of asterisks lacking brackets indicates that all pairwise comparisons within that group are significant. *$p<0.05$, **$p<0.001$ after Bonferroni correction (Wilcoxon test).