Spring 2012 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 3 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 12:30 am on Wednesday February 15. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE AT 12:30 pm ON WEDNESDAY FEBRUARY 15**. You may use a calculator, a ruler, your notes, the book, and the internet. You may take this exam in as many blocks of time as you want. Submit your electronic version before 12:30 pm (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 (no permission required since your exam is a private document).

DO NOT READ or DOWNLOAD ANY NEW PAPERS FOR THIS EXAM. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS.

-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?

average grade was 92% with 14 points added

20 pts

1) CIS major Leland Taylor ('12) has developed a web site to help students understand genome sequence assembly. Go to his web site (http://compbio.davidson.edu/phast/) and click on the popup tab at the bottom of the browser frame. Generate a phage genome sequence by having the web site assemble any particular genome but use only 10% of the reads and clean the reads as part of the assembly. Do not submit any more sequences until results from this request become available as shown by the menu at the top right of the PHAST web page. WARNING: Only assemble using 10% of the reads or it will take you a VERY long time to work on this problem.

a) Explain to a Bio111 student what an MID is, and why it needs to be removed before assembling the genome. **Your answer cannot exceed 3 sentences.**

Sequence tags that allow multiplex sequencing but individual genome separation prior to assembly. Removal is necessary to prevent false overlaps.

b) Use the Comparison Tool from the menu at the top to generate a comparison of one genome to itself. Insert a screen shot and explain what genome features you can see from the graphical output. **Three sentence maximum limit for your answer.**

Looking for comments about duplications and lack of indels

c) Find two genomes that look related to each other but do not share the same name (e.g. comparing Shanna_1 and Shanna_2 is not allowed). Take a screen shot of the alignment and then describe how your data support their relatedness. **Your typed answer cannot exceed 3 sentences.**

Lots of different possibilities. Indels, repeats, inversions, etc.

15 pts

2) To get credit for this question, you must tell me where you found your answers. The database is an important part of your answer.

a) Find the amino acid sequence for the enzyme responsible for queen formation in honeybees. Cut and past the sequence here and provide the accession number associated with your sequence.

dnmt3B as in the paper we read

```
>ACCESSION_ADH84015_honeybee
MLSEEGKLWVYWIGEARISLLNEKTQIEPFSCNLKARLTQNLNV
PRIRAIDATMQMLRKKLGGTLTKPYFTWIESNFPKNMIEMLDEIKFYPYPVKMQQRLD
HLREKNAKVTERYLLDQKRENQEKKLAEKSKDSPQKVNVDLTLLPLKEQKPGIIAWAK
IAGHNWWPAMIIDYRDCCMREPTFGCQWIMWYGDYKLSEVHHQLFLRFDKGMEKMRDY
TSNTKKHIYLVGVLQASKDYCSRLGFDTSNWTLDDAFEYFSKPNHYDYASSANTWRRE
DSVKIYDKYSARIAEKLNELKDNPNVDDQRANDINNSDDLRSAIKGEISFDSLCLKCL
RVSNDEMDIHPFFEGSLCKDCSERYKPCMFVFGNDSKCFYCTVCAASGMVIICDKEDC
PRVYCTACMKHLLCPTTYEQVLQEDPWECFLCKSRSFTTDTIVRPRANWKDKIINMFR
TSCDSNVEHLVAKHNSEKRKIRVLSLFDGLGTGLLVLLKLGFIVDAYYASEIDQDALM
VTASHFGDRILQLGNVKDITCNTIKEIAPIDLLIGGSPCNDLSLANPARLGLHDPRGT
GVLFFEYRRILKLVRKLNNERHLFWLYENVASMPSEYRLEINKHLGQEPDVIDSADFS
PQHRLRLYWHNFPIEPRLLSSQREQDVQDILTPHCQRYSLVKKIRTVTTKVNSLKQGD
GKLALKPILMKDESDSLWITELEEIFGFPRHYTDVKNLSATKRQRLIGKSWSVQTLTA
IFESLCPFFERDIVEIEG
```

b) Find DNA methyl transferases from a mammal, fish, plant and prokaryote. Provide their accession numbers.

Variety of answers. Needed to be dnmt, 3B was closest ortholog

c) Are any of these 5 protein sequences actually orthologs or are they just proteins with similar

function? Support your answer with data.
Should have done CLUSTLW or multiple BLAST alignments. Needed a way to look for conserved domains within proteins.

15 pts
3) Open the PDF of the paper showing a mutation mapped to **chromosome 7**.
a) What is the name of the human ortholog that is the subject of this paper, and what is the human chromosomal location? **Your typed answer cannot exceed 1 sentence.**
CDH2 on Chromosome 18q11.2 or .1
b) Search a human database to see if there are any alleles of this gene associated with the human orthologous behavioral condition. Provide the name of the database and how you searched for your answer along with what you found. **Your typed answer cannot exceed 3 sentences.**
Needed to search OMIM but there are no mentions of this allele and OCD.
c) Find the exact nucleotide in the dog genome and then document whether this nucleotide is conserved in humans or not. This is a challenging question, so you will have to be creative in how you document your answer. Support your answer with visual data. **Your typed answer cannot exceed 3 sentences.**
Humans have the exact base that causes doggie OCD. You should have visualized this and commented on the surprise that we have doggie OCD nucleotide.

15 pts
4) Read the PDF file called "**Paper_ette_redcated.pdf**". Explain panels a, b and c of Figure 1. You can use the text to give you some background information, but the primary point of this question is to see how well you can interpret complicated genomics figures that utilize many of the topics and results we have discussed in class. **Your typed answer to each part cannot exceed 3 sentences.**
Panel a: upper tree shows Asians and Australians more closely related, lower tree shows Europeans and Asians more closely related with later admixtures.
Panel b: PCA shows Australian more related to Aetas and Mundas speakers and less so East Asians
Panel c: no European contamination, + similar to panel b above

20 pts
5) Open the PDF file called "gut_bugs" and answer these questions.
a) Define enterotype so that a Bio111 student could understand. How many enterotypes exist? **Two sentence maximum limit for your answer.**
communities of microbes, 3
b) Circle a portion of Figure 1 that shows a group of bacteria that do not grow well with diets rich in caffeine, artificial sweetener and alcohol. Submit a screen shot of your answer and insert the screen shot here.
blue squares
c) How many enterotypes did this paper classify? How did they justify their number compared to what you found for your answer in part a? **Your typed answer cannot exceed 2 sentences.**
2 in this paper

d) Which enterotype prefers sugars over proteins? You cannot get any credit for this question unless you support your answer with data (i.e. guessing is not allowed). **Your typed answer cannot exceed 2 sentences.**

Type 2 prefers sugars

15 pts

6) Here is an abstract from a paper that summarizes some research.

a) Summarize the conclusions in **two sentences or less** so that a Bio111 student could understand.

Typically, methylation of a promoter silences the gene. In embryonic neuronal precursors, methylation flanking a promoter or between genes blocks the repressor protein polycomb from binding which causes these genes to be actively transcribed.

b) Draw a picture and insert your drawing into this Word file that visually explains the conclusions. You can draw by hand or electronically. Label as needed.

various possibilities

> DNA methylation at proximal promoters facilitates lineage restriction by silencing cell type–specific genes. However, euchromatic DNA methylation frequently occurs in regions outside promoters. The functions of such nonproximal promoter DNA methylation are unclear. Here we show that the de novo DNA methyltransferase Dnmt3a is expressed in postnatal neural stem cells (NSCs) and is required for neurogenesis. Genome-wide analysis of postnatal NSCs indicates that Dnmt3a occupies and methylates intergenic regions and gene bodies flanking proximal promoters of a large cohort of transcriptionally permissive genes, many of which encode regulators of neurogenesis. Surprisingly, Dnmt3a-dependent nonproximal promoter methylation promotes expression of these neurogenic genes by functionally antagonizing Polycomb repression. Thus, nonpromoter DNA methylation by Dnmt3a may be used for maintaining active chromatin states of genes critical for development.