Spring 2011 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 <u>not allowed discuss the test with anyone</u> until all exams are turned in at 9:30 am on Wednesday February 9. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE AT 9:30 am ON WEDNESDAY FEBRUARY 9**. You <u>may</u> 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 9:30 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.

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?

15 pts

1) Below is a diagram showing the "hypersalinity island" from the genome of *Salinibacter spp*. Answer these questions based on the figure and information you can find online, but not any research papers. Do NOT look for any papers to answer this question. You answers should document where you got your information. Include hyperlinks whenever possible to facilitate my verifying your answers.

a) What evidence can you find in credible databases that supports the claim of horizontal gene transfer (HGT) for this hypersalinity island?

b) What is biologically striking about the HGT in this particular example given what we think is the mechanism for HGT? Support your answer with data you obtain.

c) Use a genome map of the original species to identify two discrepancies/errors in this map. Limit your answer to 2 sentences or less.

> **Malcolm Campbell 2/9/11 7:21 PM**
> **Comment [1]:** GC% check is negative. -3 BLAST hits are verified. -2

> **Malcolm Campbell 2/9/11 7:23 PM**
> **Comment [2]:** Clustered and one virus infected all these species? -2

> **Malcolm Campbell 2/9/11 7:23 PM**
> **Comment [3]:** Several gene names changes. -4
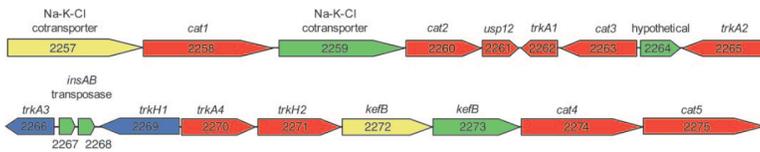
Fig. 3.   A schematic representation of the hypersalinity island identified in the genome of *Salinibacter*. Genes are color-coded with respect to their closest matches in BLAST sequence similarity searches: haloarchaea, red; cyanobacteria, green; methanogenic archaea, yellow; firmicutes, blue.

15 pts

2) You may have heard the news story in 2010 that people of Euroasian descent are genetically 1-4% Neandertal. The figure to the right shows 2 scenarios (ignore numbers 1 & 4) of when genetic information could have been exchanged. Describe what you would see at the DNA level when comparing the 5 populations along the top of this tree if scenario 2 were correct or if scenario 3 were correct. In your description, use either cartoon figures or DNA sequence examples you make up to illustrate your answer. Your answer should include answers for both scenarios 2 & 3, not just one scenario.
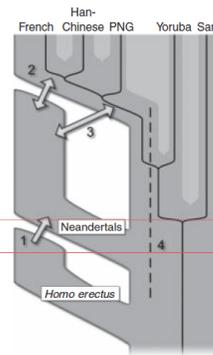
> **Malcolm Campbell 2/9/11 8:24 PM**
> **Comment [4]:** San and Yoruba would never have any Neandertal DNA. -5

> **Malcolm Campbell 2/9/11 8:24 PM**
> **Comment [5]:** This would show sequence only in French. -5

> **Malcolm Campbell 2/9/11 8:24 PM**
> **Comment [6]:** This would show sequences in PNG, Han and French with some differences between them probably. -5

15 pts

3) One of the great things about genomics is you can ask historical questions even if the organism in question is extinct. For example, investigators have wanted to know how the wooly mammoth adapted to colder climate. We knew that longer fur, shorter tail and smaller ears all helped, but how could their hemoglobin function given that in most mammals, O_2 binds more tightly to hemoglobin in colder temperatures. A group of genomicists sequenced the mammoth hemoglobin A (HBA) and the chimeric hemoglobin B/D (HBB/HBD) genes and the resulting phylogenetic trees are shown below.

Figure legend: Branch lengths are not proportional to geologic time. Ancestral nucleotide and amino acid residues are shown above, and derived nucleotide and amino acid residues are shown below the

2

exons. The numbers above and the letters to the right of the vertical lines denote the amino acid residue, whereas the numbers below and the letters to the left of each vertical line indicate nucleotide position relative to the ATG initiation codon. Thick vertical lines with bold characters indicate nonsynonymous substitutions, and thin vertical lines represent synonymous substitutions, with red, green and blue characters and bars representing replacements at codon positions 1, 2 and 3, respectively.
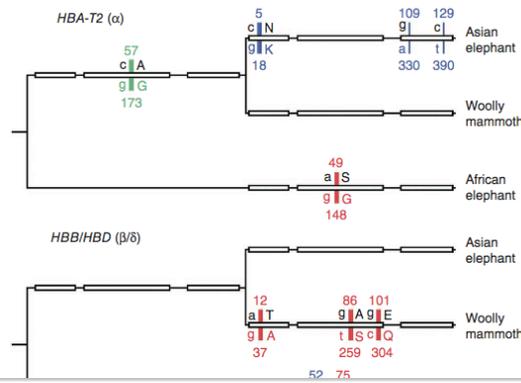
a) Interpret these phylogenetic trees.
b) How many changes does each of the 3 terminal **proteins** have (answer for both trees separately) compared to the root?
b) Align the amino acids for the 3 terminal proteins of HBB/HBD and verify the mutations shown in the figure. You have to show me a screen shot of the alignments and tell me how you did the alignments. Your alignments may or may not agree with the figure.
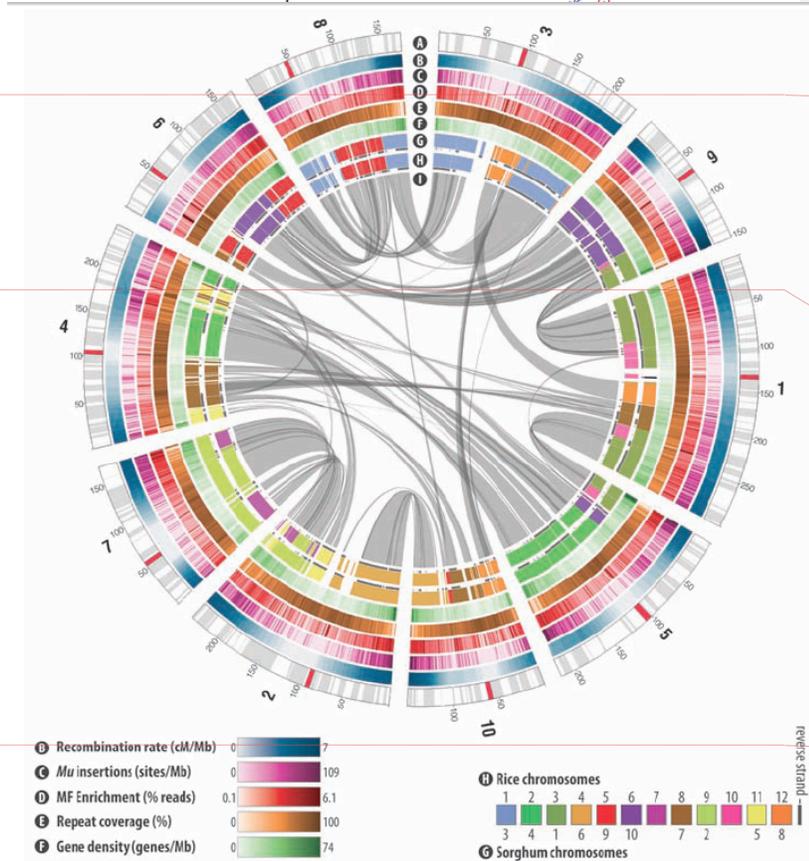
15 pts

4) The colorful circles on the right illustrate various features of the maize genome. Note the legend at the bottom left corner. Summarize in 3 sentences or less for each prompt below what you can determine for each of following layers:

B:

Malcolm Campbell 2/10/11 12:00 PM
Comment [7]: HBA – Asian and Wooly have the green in common, African is different.
HBB/D Asian is like the root. Wooly and African are each different. Hard to see how the trees were drawn this way. Why not Asian at the base and African and Wooly branching off from it?
3 pts

Malcolm Campbell 2/10/11 12:00 PM
Comment [8]: Top:
Asian = 2
Wooly = 1
African = 1
Bottom:
Asian = 0
Wooly = 3
African = 1
6 pts

Malcolm Campbell 2/10/11 12:01 PM
Comment [9]: various correct answers must be amino acid
6 pts for aligning Asian to Wooly and to African

Malcolm Campbell 2/10/11 2:43 PM
Comment [10]: recombination rates lowest near centromeres and highest near telomeres.
3 pts

E:

F:

G&H:

I:

15 pts

5) Using the Venn diagram to the right, answer these questions:

a) How many are shared by all monocots but not dicots? 2077

b) How many are shared by all 4 plants? 8494

c) Which plant has the largest unique number? Speculate what aspect of its biology led to this outcome. Rice = 1110 = varied habitat.

d) Which plant has the next largest unique number? Speculate why this is the case. 1058 = dicot vs. monocots

e) What do these numbers represent? What is missing from this figure?
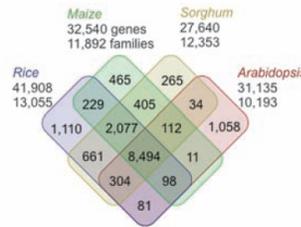
Gene families allows you to determine paralogs/isoforms. One gene family is missing from the 8494 number.

Maize
32,540 genes
11,892 families

Sorghum
27,640
12,353

Rice
41,908
13,055

Arabidopsis
31,135
10,193

465 265
229 405 34
1,110 2,077 112 1,058
661 8,494 11
304 98
81

15 pts

6) To the right is a figure showing synteny between chicken, frog and human genomes. Support your answers by explaining your logic. Using your understanding of genomics, speculate what is indicated by:

a) the black triangles above the image

b) the gray hashed regions on the human chromosome

c) the white shapes on the black background of the chicken genome.

A

Frog   Chicken

23   3
21   8   8
VI   IX
IV   VII   VII

0   50   100   150   200   250
Position on Human Chromosome 1 (Mb)
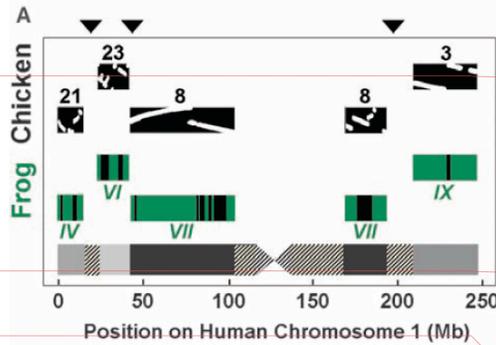
10 pts

7) This question is a stretch and very challenging. I want to see if you guys can open your minds and listen to what the data are trying to tell you. This one requires you to think and use logic to deduce the answer.

The graph on the right measures the ratio between centimorgans and DNA length (cM/Mb) in one generation of a human family (2 parents and 2 children). Given that the length of a chromosome is relatively stable, how can the cM/Mb ratio vary

window size (bp)
10^5
10^4
10^3
10^2

0.01   0.1   1   10
recombination rate (cM/Mb)

over 4 orders of magnitude? Explain the implications of this graph with respect to the human genome. FYI: this graph is for one particular chromosome, but a similar graph would be produced for all chromosomes. **Keep your answer to 4 sentences or less.**