Spring 2017 Genomics Exam #1
Genomic Sequences & Variations

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. You are not allowed discuss the test with anyone until all exams are turned in at 2:30 pm on Wednesday February 15. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 2:30 pm ON WEDNESDAY FEBRUARY 10**. 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 2:30 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 storage). 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 NEW 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.

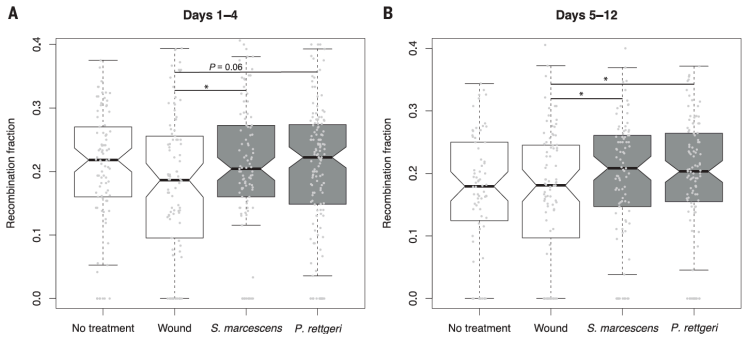
Name (please type):

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

How long did this exam take you to complete?

10 pts

1) The offspring of some experimental fruit flies were measured for recombination frequency in their chromosomes under the four conditions shown here: no treatment; sterilely wounded by investigators; infected by two bacterial species as indicated. * = p < 0.05. Whiskers = most extreme data points, black line = median, boxes = second and third quartiles. Days indicate post manipulation time.



a) Interpret the experimental results in panels A and B based on the presented data.

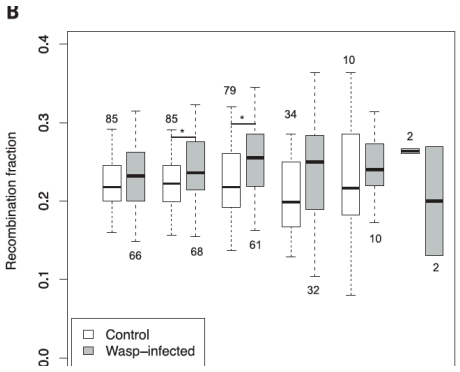
b) What are the evolutionary implications of these data?

10 pts

2) Continuing the fly research, here the recombination rates are grouped into 2 day bins post manipulation. * = p < 0.05 for paired data within a bin. Numbers above and below the whiskers represent the number of replicates for control or treatment.

a) Interpret these experimental results based on the presented data.

b) What molecular mechanistic implications can you speculate about using these data?

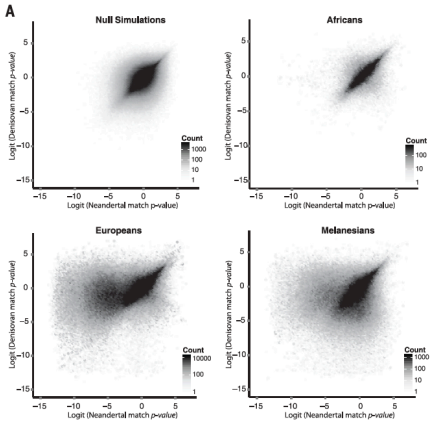


10 pts

3) A group of genomicists wanted to find the origins of ancient hominid DNA present in modern human genomes. They sampled Esan people from Nigeria, northern Europeans, and people from Melanesia. Y-axis measures p-values for sequence matches between modern and Denisovan DNA; X-axis measures p-values for sequence matches between modern and Neandertal DNA.

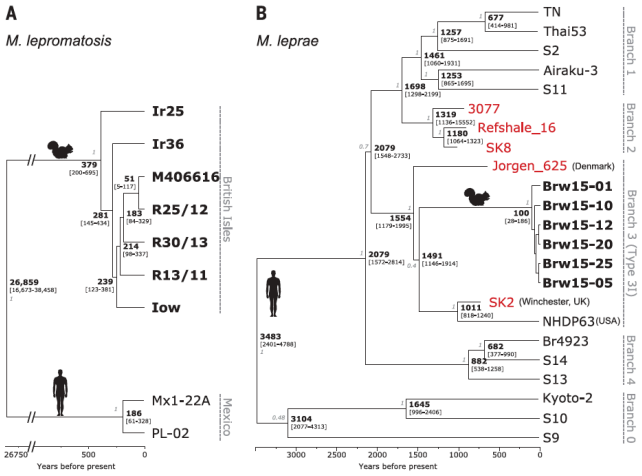
a) Interpret all four panels from this figure.

b) Address the evolutionary significance of this finding given that the investigators compared ancient and modern DNA from about 160 geographically diverse modern populations.



15 pts

4) Two species of leprosy-causing bacterial exist in the world. Panel A shows the evolutionary analysis of two populations, one from the UK and one from Mexico. The small black silhouettes show which host each strain came from (human or squirrel). Panel B shows a similar analysis for the other leprosy species. Red text indicates ancient samples, bold indicates squirrel samples.



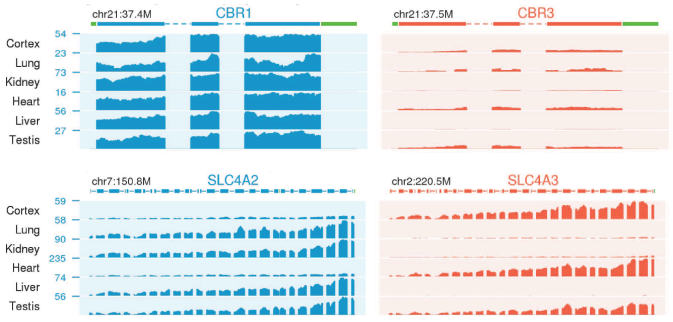
a) Interpret the evolutionary relationships between the two different sources of bacterial genomes for panels A and B separately. Include a time scale in your answers.

b) Which human sample is the MRCA for the squirrel samples in panel B, based on the data?

c) What can you deduce about leprosy in British squirrels over the last 2,000 years given the investigators only sequenced squirrel samples from live animals?

15 pts

5) The human genome contains many paralogs, as shown in these examples. Gene structures are shown above with histograms below showing transcriptome data.



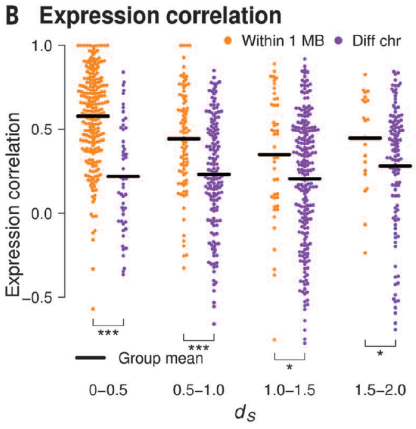
a) Interpret the data for the top pair (CBR1 & 3).

b) Interpret the data for the bottom pair (SLC4A2 & 3).

c) Use these data to generate two hypotheses of what can happen to paralogs after gene duplication.

15 pts

6) Continuing with human paralogs, here you can see investigators measured the correlation of transcription levels across many tissues for paralogs on the same chromosome or on different chromosomes, as indicated by orange and purple dots. * = $p < 0.05$; *** = $p < 0.001$. d_s is a proxy for time since the gene duplication event that produced paralogs. Bigger numbers means more years since duplication.



a) Interpret the figure for all four d_s categories.

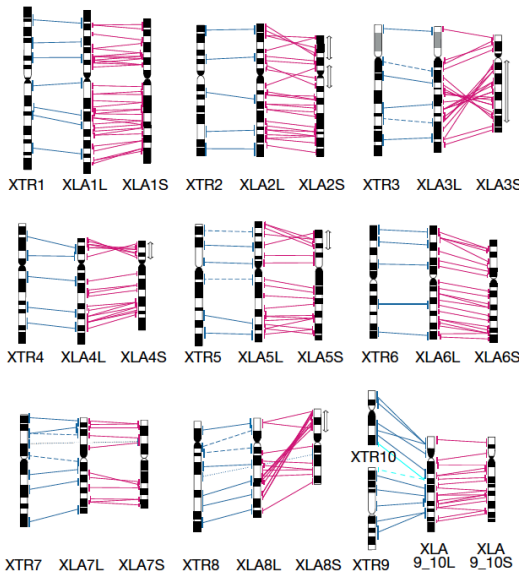
b) Which of the two scenarios in Question #5a and 5b best explains what is happening to paralogs on the same chromosomes? Support your answer using data from both questions 5 and 6.

15 pts

7) *Xenopus laevis* (XLA) is a tetraploid frog whereas *X. tropicalis* (XTR) is a diploid; both frogs have had their genomes sequenced. The diagram shows 9 XTR chromosomes and how they map onto XLA chromosomes (L = large and S = small). Blue lines show positional relationships between conserved orthologs for both species (solid), only with XLAS (dotted) or only XLAL (dashed). Double headed arrows indicate inversions. Magenta lines show relationships between XLA paralogs.

a) For XTR chromosomes 1 – 9, indicate which XLA chromosome is more similar to XTR.

b) Integrate the data from Questions 5 – 7 to speculate what is happening to the XLA paralogs.



10 pts

8) Using the ExAC database, the figure to the right shows higher Z scores for SNPs that are selected against. Gray = synonymous SNPs, orange = missense SNPs, red/rust = nonsense SNPs.

a) Interpret the data based on this figure.

b) ExAC found one SNP for every 8 bases. What is the evolutionary implication of the data in the figure given the assumption that all SNP mutations are equally likely to occur?