

**Fall 2008 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 much less time than 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. You will have to answer 6 Discovery Questions and 2 original questions. I expect the original questions will take you more time than the Discovery Questions. You are not allowed discuss the test with anyone until all exams are turned in at 11:30 am on Wednesday October 1. **PAPER COPIES OF YOUR EXAM ANSWERS ARE DUE AT CLASS TIME ON WEDNESDAY OCTOBER 1.** You may use a calculator, a ruler, your notes, the book, and the internet. You may take it in as many blocks of time as you want. Submit your paper and electronic version before 11:30 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 ONLY ON THE FIGURES PROVIDED IN THIS EXAM, 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 print):

Write out the full pledge and sign:

"On my honor I have neither given nor received unauthorized information regarding this work, I have followed and will continue to observe all regulations regarding it, and I am unaware of any violation of the Honor Code by others."

How long did this exam take you to complete (excluding typing)?

For this part of your exam, answer these 8 Discovery Questions. Be sure your answers are supported with data and spell out your logic. Screen shots and URLs are good ways to validate your answers.

I cannot post answers to these since some faculty use them for graded exercises. I would be happy to answer any questions you have.

6 Pts Chapter 2 #69: The options have changed, so you will have to figure out what variations you need to do to answer the BLASTx question. Don't worry about Genefinder.

6 Pts Chapter 2 #71: This one requires some hunting to track down.

6 Pts Chapter 2 #72: tell me how you reached your answer.

6 Pts Chapter 2 #73

6 Pts Chapter 2 #74: The link is not working so you will need to hunt this page down. Get a screenshot to show me that you found it.

8 Pts Chapter 3 #16 Describe the significance of the differences you see. Use screenshots to support your answer.

6 Pts Chapter 3 #22: Read carefully as you go to follow this one.

6 Pts Chapter 3 #71: To get full credit, you need to tell me how your reached your answer to be sure. Also, describe any interesting aspects of the genes you find.

30 Pts

1) This question is one that requires some good detective work.

a. What is the original source of this sequence?

LYFFFLCTMGLVMSXXXXAFQCYHYTLTNSFSCREMXXXCICTLDPEDPIARMFVF

Give me the Latin and common name of the species as well as the name of the protein which happens to be related to a protein we have already studied.

b. Retrieve 6 orthologs of this query sequence. List the common names for species you have chosen.

c. Go to this site (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) and generate a phylogentic tree showing the relationship between these sequences. Be sure your tree uses the common names for labels. Interpret the tree as fully as you can.

d. Generate a WebLogo for 30 highly conserved amino acids. Tell me what site you used to generate your logo, which amino acids you chose, and why.

e. Do you think the human ortholog is an integral membrane protein? Support your answer with data other than reading it somewhere. Use a bioinformatics tool to independently validate your answer.

f. Go to this site (<http://frodo.wi.mit.edu/>) and design PCR primers that will amplify a piece of DNA between 500 and 600 base pairs using this human DNA as template. Use screenshots to demonstrate you have a good pair of PCR primers.

>Zebrafish sarcospan

MGLGMNPKVGGNQSSEKKAKAETGGPVPEEAQKCCGCRFPLLVALLQLLLGIAVAFAVAF
 LMLAISSSLLARETPHWAGIIMCVVSLGFLVFCITRVPDERALLQFVIKLLYFFLCTM
 GLVMSVVVIAFQCYHYTLTNSFSCREMREDCICTLDPEDPIARMFVFSVSDCSDITST
 LPMYYLLQVLLNLAQAIIVCLLGAFLIWKHRYQVFFAGLQGTGSPAAQHWQKA

>*Tetraodon nigroviridi*

MGQKKDKESRSEKREGSPEAEDGHKCRFCRFPVVALLQLLLGVAVTVVAFMLAISPS
 LLARETPHWAG
 IILCLVSVLGFILYCITYLPDERTSAQFICEVLYFILCSIGLVLSVLVVAFAAGHHYSQA
 SGFSCGPARGD
 CVCTLSQEDPIARTFTYEGVGDCAITGTLTLYFLVQIVLNLAQALVCAAGAFIMWKHR
 YQVFFAGLQI

>*Gallus gallus*

MGKKEKKVQHNAAHGEEQGTDSGHTQEPAMKKKKKKKSKGEPKAGWEEESHTECCWCRFP
 LLCALLQLAFG
 VAVTVLGLFMAGISSSLPVRDTPYWAGIIVCVVSLVGFVMLCISYQPDEKTCVQFTVKL
 MYFLLSALALV
 LCVVAFAFAAHHYLQMTKFTCDTVLESCQCKLDTVDPLSRTFVYQDAADCSSVTSMLSL
 YLILQMVNLNLL
 AALVCLSACFIMWKHRYQVFYVGARFYPLTTTECQQQKV

>*Xenopus laevis*

MAAGKDKGNESAPEPKKEAKVGREESHNCCGCRFPLLIALLQSLGVSVTVLSFIMAAS
 CPSLLPRDTPY
 WAGICVSAVAVLGLILLCLPYQPDEKTMCFVLKLLYFLLSALGLIICVTAVAFAYHH
 SYITKFTCHMA
 TDWCQCTMDSSDPLSRTFQYQNVSDCDSITGTMKLFVLLQMALNLLLALVCLASCFVMW
 KDRYQVFYVGQ
 WFPGPAPNVARQQKV

>*Oryctolagus cuniculus*

MGKDRQPRGQQRQGDAAAGPDDPGPKKGAGTREQRGEEEAQTCCGCRFPLLLALLQLALG
 VAVTVVGFLMA
 SVSSSLLVRATPYWAGIIVCVVAYLGLFMLCVSYQVDERTCIQFSMKLLYFVLSALGLV
 VCVLAVAFAAH
 HYSLLTHLTCENAPDSCQCKLPSSEPLSRTFVYRDVTDCTSITGTFQVFLLVQMVNLNV
 CGLVCLVACFV
 MWKHRYQVFYVGVRCPLSASEGQQQKV

>*Macaca mulatta*

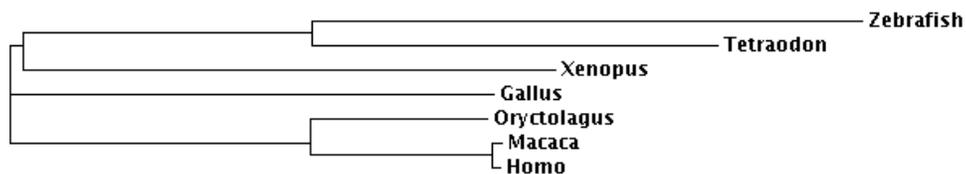
MGKNKQPRGQQRQGGPPAADAAGPDDMEPKKGTGAPKECGEEEPRTCCGCRFPLLLALL
 QLALGIAVTVV
 GFLMASISSSLLVRDTPFWAGIIVCLVAYLGLFMLCVSYQVDERTCIQFSMKLLYFLLS
 ALGLTVCVLAV

AFAAHHYSQLTQFTCETTLDSCQCKLPSSEPLSRTFVYRDVTDCTSVTGTGTFKLFLLIQM
 ILNLVCGLVCL
 LACFVMWKHRYQVFYVGVMRSLTASEGPQOKI

>Homo sapiens

MGKNKQPRGQQRQGGPPAADAAGPDDMEPKKGTGAPKECGEEEPRTC CGCRFP LLLALL
 QLALGIAVTVV
 GFLMASISSLLVRDTPFWAGIIVCLVAYLGLFMLCVSYQVDERTCIQFSMKLLYFLLS
 ALGLTVCVLAV
 AFAAHHYSQLTQFTCETTLDSCQCKLPSSEPLSRTFVYRDVTDCTSVTGTGTFKLFLLIQM
 ILNLVCGLVCL
 LACFVMWKHRYQVFYVGVRICSLTASEGPQOKI

Phylogram



Notice that by submitting the entire sequences, the tree more closely matches what we think of for evolution. Submitting partial sequence gave different results. Also, when searching for orthologs, you would be better served by submitting the full zebrafish protein sequence rather than just the partial sequence with Xs. Finally, regardless how you got your tree, you should have commented on how the PROTEINS are related, not the species.

Multiple Logos were acceptable, but they had to show variations. You should have chosen your sequence from the entire protein, not just the fragment I gave you. Also, you should have used your alignment to find the highly conserved amino acids.

You should have used Kyte-Doolittle to determine sarcospan is a probable membrane protein with a screen shot showing 4 TM domains. You could site the textbook for validation.

20 Pts.

- 2) Like the last question, this one requires that you think creatively to find the answers.
- What species is associated with this identifier (ATCC 43049), and where was it found in nature?

Everyone found this.

b. How many DNA segments does this organism have? Support your answer by giving me the URL for your source of information.

This was to be collected from the web site where you found the species DNA information.

c. How many of these segments contain RNA genes? How many chromosomes does this species have? Explain your answer.

I really wanted you to discuss the controversy about chromosome vs. plasmid. Since 2 plasmids have RNA genes, and since the sizes vary a lot, it is worth noting that the number of chromosomes is a debatable figure.

d. Use this genome's data to demonstrate it probably has a virus that can infect it. Support your answer with the name and accession number of a gene that is one of only 36 proteins encoded on a single DNA segment.

Everyone found this.