Fall 2001 Genomics Exam #1
Genomic Medicine and Sequencing Tools

There is no time limit on this test, though I have tried to design one that you should be able to complete within 2.5 hours, except for typing. There are three pages for this test, including this cover sheet. You are not allowed discuss the test with anyone until all exams are turned in at 9:30 am on Friday September 21. **EXAMS ARE DUE AT CLASS TIME ON FRIDAY SEPTEMBER 21**. You may use a calculator, a ruler, your notes, the book and the internet. However, you are not allowed to obtain and read journal articles as a part of your investigations. These questions are taken from the research literature and I do not want you to simply find the papers and read the answers. This is the Honor Code at its finest.

The **answers to the questions must be typed on a separate sheet of paper** unless the question specifically says to write the answer in the space provided. If you do not write your answers in the appropriate location, I may not find them. You may want 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.

-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:

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

The answers listed below are not complete and full, but list the main points.

20 points

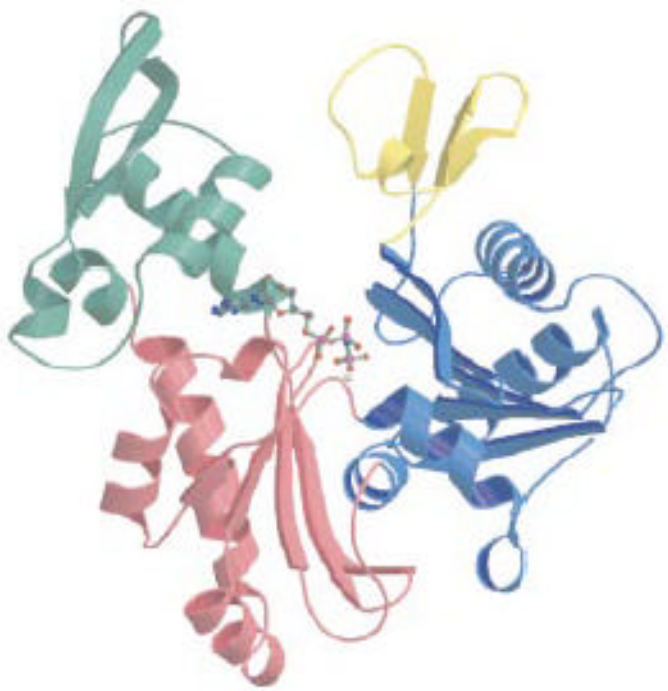
1) Sequence Number 1:
mfgrdigidl gtanvliyvk gkgivlneps
DEFINITION cell shape determining protein (MreB-like protein) [Bacillus halodurans].
ACCESSION NP_244606

Sequence Number 2:
MDSEVAALVIDNGSGMCKAGFAGDD
DEFINITION actin [Saccharomyces cerevisiae].
ACCESSION CAA24598

a) From what protein did each of these sequences come?
b) What are their accession numbers?
c) Compare their overall similarities. How much sequence conservation is there?
None

The first protein was recently crystallized and its X-ray structure determined to be very similar to the second protein. The authors conclude that the first protein is a functional "homolog" of the second one.

Here is a photo of the first one's structure. The colors are not PDB standard colors.



d) Compare the photo above with this PDB file (http://www.bio.davidson.edu/people/macampbell/genome_exam/exam1.pdb) and tell me if you think these two proteins are orthologs, "homologs" or neither. Explain your answer.

Structurally very similar, including an ADP bound in the cleft. Convergent evolution? Funcional homologs but not sequence homologs.

15 points

2) Go to this URL http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl and compare the Genome v. Genome Protein Hits for *Mycoplasma genitalium* G-37 and *Ureaplasma urealyticum* serovar 3.

a) How many genes do they have in common?

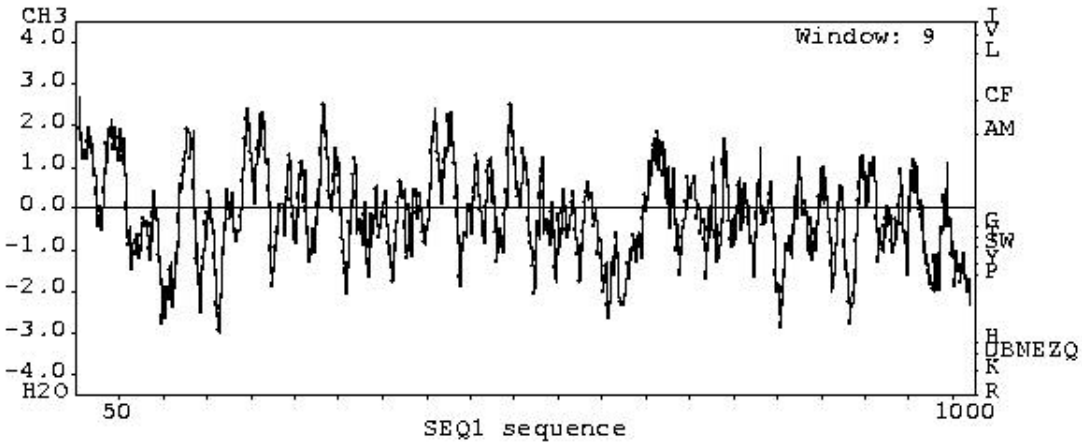
89 genes in *M. genitalium* match *U. urealyticum* but **94** genes in *U. urealyticum* match *M. genitalium*.

b) Find one area of highest block of conservation in the *Ureaplasma urealyticum* serovar 3 genome and click on this part of the green circle. Tell me what kind of protein is in this region.
c) Would you expect to find this protein highly conserved? Explain your answer.
Ribosomal proteins. Not a surprise.

20 points
3) Tell me everything you can about this sequence. It is **very important** that you list all the sites you used to answer this question. For each site, tell me what you learned.
http://www.bio.davidson.edu/people/macampbell/genome_exam/exam2.html
Chlamydomonas ezy2 sequence. Other species had E values too big to be of value.

Translation of ORF

```
>lcl|Sequence 1 ORF:183..3260 Frame +3
MAVACAVAVRPLVQVAVASAVSTAAPASSKPAVKLAASAVSAVALTTVSVSAGLLATTAVEDPRFHAADC
QSRSADASASCEDLQPSTSTCTSAVRDANRPTRRVRRSGSKAQRRGSTTLTASVPSMAAAVVLPPKIALR
RRHRLRLRAGHSATAAATDKTPREQPDKPAALPEDLLPADATSTSSTGKISSAAVCCGLLAHCSAAQLHA
ILCGLVQAVASSSVKGNNRKLLLGSKLRKLLEGVGVAPANGKAYTAADVAALSGPKLERLRATLKSQPGL
LLWFLLFTAPAKLQALQAALLPGGAGDRSFEEWRAAIDAVAGSGHEQLAAAQEVRGRQSACVEGSTAGNT
ATTATITTTNNNPASHGGVYTALTGTEVTGKKPAALPEDLLPADATSTSSTGKISSAAVCCGLLAHCSAA
QLHAILCGLVQAVASSSVKGNNRKLLLGSKLRKLLEGVGVAPANGKAYTAADVAALSGPKLERLRATLKS
QPGLLLWFLLFTAPAKLQALQAALLPGGAGDRSFEEWRAAIDAVAGSGHEQLAAAQEVRGRQSACVEGST
AGNTATTATITTTNNNPASHGGVYTALTGTEVTGKAAANKDLSRTRTTSHRNRCVSESGSTRNKSRSSSS
RSSSTHSVEYAEPKAGCSQPAATVPGCVPEIISAAIPPLAPLALHIRRAIVKELLEARPPGWNTFLYSWL
QAAGLSEFLPANGTCRMYMADRKQLVLRVGAMREEQVDAFLTCMCKAHGHSTWLARYLHMLGPEVSQLLS
QGRYSDELLAALRAAGQKTLADAVMEHFWGRDPDPEDSEAGEMDVKPWAERLGLLRFDMLAEQLRLPPNA
DGSVKNFSNGLVFKVDPLEVWSKYTDGEPSAGALSGMRATDKEARDKQVKQLRGVPLLYLWRIGGRVVYV
GMSGGWVKGRRIARYLAEGPGFSESSKMLPWLTAIDEGKEIELRVITLEGLKALEGMSEGMSEEEVQKKV
QKKVKELEKHFLCHVDCPCNKVNNGSYRVETPRQASWTNSRRSTR*
```



PDB, Conserved Domain gave no information. PubMed described how ezy2 is involved in organelle inheritance in Chlamydomonas.

15 points

4) This sequence was obtained from a specimen collected in 1847.

AATTTCTCCAACAAAATCTTAACCTGGAATAGACATATTTGCTAATACATAAATAAA

This sequence was collected from a specimen in 1994

AATTTCTCCAACAAAACTACTTGAACCTGGAATAGACATATTTGCTAATACATAAATAAA

a) Are these two sequences related?
b) Describe the amount of relatedness that you find.
c) What was the source of the first sample dating back over 150 years ago?
d) Find a recent paper that discussed the significance of these sequences. What does the abstract say about these sequences.

Irish potato blight. 93+%.
Ristaino et al. 2001. Nature. Vol. 411: 644.
Strain Ib probably did not cause blight as had been assumed. This haplotype not found in historical DNA sample.

15 points

5) Read this short paper on fish and briefly evaluate it from a genomic perspective. Your evaluation should focus on the conclusions drawn from the data and not the data themselves. Answers over 1 page will be too wordy, so don't go for length on this one.

Devlin *et al.* 2001. *Nature* Vol. 409: 781-782.
This is a simplistic view of genomics. Bovine growth hormone, one genotype v. many (wild). Complex circuitry. Many good points to be made here.

15 points

6) Go to this URL http://devnull.lbl.gov:8888/alt/search.html and perform a search of all listed SP fields using this term "ATC2_HUMAN".
a) How many functionally different proteins can this gene produce?
b) In which tissues?
c) Are their any functional differences for these proteins?
d) Based on what you found, are similar traits found in other species or is this limited to humans? Support your answer with data.

Follow the links to "NiceProt" http://www.expasy.ch/cgi-bin/niceprot.pl?P16615

* ***FUNCTION***: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE TRANSLOCATION OF CALCIUM

FROM THE CYTOSOL TO THE SARCOPLASMIC RETICULUM LUMEN. ISOFORM SERCA2A IS INVOLVED IN THE REGULATION OF THE CONTRACTION/RELAXATION CYCLE.

* ***CATALYTIC ACTIVITY***: ATP + H_2O + CA^{2+}(IN) = ADP + PHOSPHATE + CA^{2+}(OUT).
* ***ENZYME REGULATION***: REVERSIBLY INHIBITED BY PHOSPHOLAMBAN (PLN) AT LOW CALCIUM CONCENTRATIONS. DEPHOSPHORYLATED PLN DECREASES THE APPARENT AFFINITY OF THE ATPASE FOR CALCIUM. THIS INHIBITION IS REGULATED BY THE PHOSPHORYLATION OF PLN *(BY SIMILARITY)*.
* ***SUBUNIT***: ASSOCIATED WITH PHOSPHOLAMBAN (PLN) *(BY SIMILARITY)*.
* ***SUBCELLULAR LOCATION***: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC AND ENDOPLASMIC RETICULUM.
* ***ALTERNATIVE PRODUCTS***: 2 ISOFORMS; SERCA2A/ATP2A2A/CLASS 1/HK2 AND SERCA2B/ATP2A2B/CLASS 2-4/HK1 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING. SERCA2 TRANSCRIPTS DIFFER ONLY IN THEIR 3'UTR REGION AND ARE EXPRESSED IN A TISSUE-SPECIFIC MANNER. SERCA2A IS A CARDIAC/SLOW TWITCH, MUSCLE SPECIFIC ISOFORM AND SERCA2B IS A UBIQUITOUS HOUSEKEEPING ISOFORM. SERCA2A HAS A LOWER AFFINITY FOR CALCIUM AND A HIGHER CATALYTIC TURNOVER RATE.
* ***TISSUE SPECIFICITY***: ISOFORM SERCA2A IS HIGHLY EXPRESSED IN HEART AND SLOW TWITCH SKELETAL MUSCLE. ISOFORM SERCA2B IS WIDELY EXPRESSED, IN SMOOTH MUSCLE AND NONMUSCLE TISSUES SUCH AS IN ADULT SKIN EPIDERMIS.
* ***DISEASE***: DEFECTS IN ATP2A2 ARE THE CAUSE OF DARIER'S DISEASE (DD) (ALSO KNOWN AS DARIER-

WHITE DISEASE; DAR). DD IS AN AUTOSOMAL DOMINANTLY INHERITED SKIN DISORDER CHARACTERIZED BY LOSS OF ADHESION BETWEEN EPIDERMAL CELLS (ACANTHOLYSIS) AND ABNORMAL KERATINIZATION. PATIENTS WITH MILD DISEASE MAY HAVE NO MORE THAN A FEW SCATTERED KERATOTIC PAPULES OR SUBTLE NAIL CHANGES, WHEREAS THOSE WITH SEVERE DISEASE ARE HANDICAPPED BY WIDESPREAD MALODOROUS KERATOTIC PLAQUES. IN A FEW FAMILIES, NEUROPSYCHIATRIC ABNORMALITIES SUCH AS MILD MENTAL RETARDATION, SCHIZOPHRENIA, BIPOLAR DISORDER AND EPILEPSY HAVE BEEN REPORTED. STRESS, UV EXPOSURE, HEAT, SWEAT, FRICTION, AND ORAL CONTRACEPTION EXACERBATE DISEASE SYMPTOMS. PREVALENCE HAS BEEN ESTIMATED AT 1 IN 50000.
*