

Fall 2005 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 work hard to design challenging tests that continue your learning and hopefully will stimulate you too. 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, and 5 questions for this test. You are not allowed discuss the test with anyone until all exams are turned in at 11:30 am on Friday September 30. **EXAMS ARE DUE AT CLASS TIME ON FRIDAY SEPTEMBER 30.** 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.**

You may want to use some of the resources on this page <<http://bioinformatics.org/sms/>> but you may not need to. Just wanted to supply everyone with a common suite of tools.

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 (by typing a second time and signing paper version):

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

20 pts.

- 1) Access the GRAMENE Genome Browser at this web site (<http://www.gramene.org/>).
 - a. Consider the synteny of rice and maize and make an evaluation of the large-scale genomic changes that have occurred when comparing these two species. When you think you have a good sense of what has happened, use a screen shot of the most extreme example of what you observed to support your conclusions.
 - b. Now search for Histone H3. Tell me where this gene is encoded in the Rice Japonica genome. Use a screen shot to support your findings.
 - c. Use data to convince me whether histone h3 is highly transcribed or not.
 - d. Is this gene similarly expressed in other grain plants? How do you know?
 - e. What strand and which reading frame encodes the h3 protein you have selected? Explain how you found your answer and provide a screen shot of the DNA and protein sequences as shown in GRAMENE that validates your findings.

20 pts.

- 2) Start with this sequence and answer the following questions:

```
CCTAGTCTCCCTCCTCTTCGTCATGCTGCGCTACATGTACCCGGCACAAGGGCACGTACCACACCAATGAGGCCAAGG
GCACGGAGTTTGCTGAGAGTGCAGATGCAGCCCTGCAGGGAGACCCTGCCCTCCAAGATGCTGGTGATAGCAGCAGA
AAGGAGTACTTTATTTGAGGGACAACAGACTTCACTTCCCTGAATGCCTCCCCATCTCCATCAGGAAAAATACACC
CCATCGCCCAGCACCCCTGCTGATACCACCAGACAGAGAGAGAGACTTGATTCTTCCCAGATAGCCACCTGGA
AACACTAGGTGCCTGCCAGGGAGGAACGGAGGAGGACTCGCGCTACAAGAG
```

- a. What gene is this?
- b. Provide me with the protein sequence.
- c. List the protein's functions and your source.
- d. Are there any STS markers for this gene? Support your answer with data.
- e. Are there any splice variants? Support your answer with data.
- f. Find the percent identity in the protein sequence of this gene with its chimp ortholog. Show data to support your findings.
- g. Tell me the percent identity for human and chimp proteins granulysin, protamine, and semenogelin. Use a table to show your results.
- h. Human and chimp have 29% of their orthologous proteins 100% conserved and the average number of amino acid changes is 2 per protein. Propose an hypothesis to explain these genome-wide numbers with the numbers you found in f. and g. above.

20 pts.

- 3) Open the attached file called worksheet.doc.
 - a. Find the largest open reading frame. In which frame is the largest ORF? How many amino acids? What is the predicted molecular weight? What conserved domains are there, if any? Show data for each of these answers.

- b. Is this the *wt* reference sequence or a mutant DNA sequence? Show data to support your answer.
- c. Given this allele of the gene/cDNA, explain how it could lead to a disease in humans. Use a Genome Browser and consider the setting “Stanf Meth” to support your hypothesis with data.

20 pts.

- 4) Find out some information about genome for the bacterium *Helicobacter pylori* J99 to answer these questions:
- How big is the genome?
 - What is the coding density?
 - What is the overall %GC?
 - Find a gene with a significantly different %GC. Tell me the gene, its %GC, and provide me with the DNA sequence.
 - Identify the origin of replication and support your answer with two independent types of supporting data.
 - What percentage of *H. pylori*'s genes are considered to be essential? Provide me with the numbers you used to calculate your answer and how you obtained these numbers.

20 pts.

- 5) I have just read an interesting paper about RNAi in mouse. See if you can make sense of these clues and data.
- Find the DNA sequences for NM_021476 and D530007L20.
 - Find their chromosomal positions.
 - Look at Figure 1.

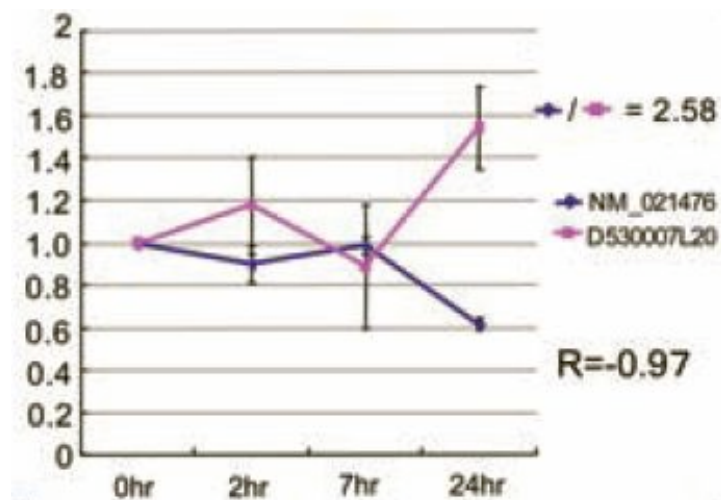


Figure 1. Time-course analysis of S/AS (sense/antisense) pairs. Expression of S/AS RNA pairs was verified by reverse transcription polymerase chain reaction over 24 hours after activation of macrophages with LPS. R, correlation coefficient. y axis, relative expression; blue/pink symbols ratio, actual expression

levels at time 0 hours.

- d. Use NCBI tools to determine if the sequences you retrieved in “a” are in fact sense antisense sequences. Show data to support your findings.
- e. Do these two coding segments have the same codon bias? Support your answer with data and interpret the implication of your findings.
- f. Now look at the file called “SOM Fig5.pdf” and try to make sense of all the data you have collected so far. Summarize your conclusions based on all the data you have, both from your own research and from this publication.