

Fall 2009 Genomics Exam #1
Genomic Sequences ANSWER KEY

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 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. There are no Discovery Questions on this exam. 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 SEPTEMBER 30.** You may 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 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 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):

Answer Key

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

"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)?

Answers here are truncated and are intended to help students understand the key points I was looking for. I would be happy to discuss any concerns you have about your answers and my grading.

Briefly read over the paper about Neanderthal MC1R gene and then answer these questions. You will NOT need to read everything in the paper first. I want you to skim the paper and then read in detail as you need to.

6 pts.

1) Find the human MC1R ortholog and describe what makes the human ortholog unusual. Support your answer with data and the source for your data.

no intron

long 5' UTR

8 pts.

2) Align the human and Neanderthal sequences to verify the differences as reported in this paper. You must show me your alignment with a screen shot. You must also explain how the DNA changes produced the amino acid changes.

```
>[ref|NC_000016.9] Homo sapiens chromosome 16, GRCh37 primary reference assembly
Length=90354753

Score = 159 bits (86), Expect = 4e-39
Identities = 88/89 (98%), Gaps = 0/89 (0%)
Strand=Plus/Plus

Query 1          TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCCTCATCTACGCCTTCCACAGC 60
                |||
Sbjct 89986513   TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCCTCATCTACGCCTTCCACAGC 89986572

Query 61         CAGGAGCTCCGCGGGACGCTCAAGGAGGT 89
                |||
Sbjct 89986573   CAGGAGCTCCGCAGGACGCTCAAGGAGGT 89986601
```

GGG to AGG is the codon change

12 pts (4 pts each)

3) Translate the human allele and paste the amino acid sequence below. Tell me the isoelectric point and the molecular weight of this protein.

Use ORF finder

Several tools online allow MW and pI calculation.

8 pts

4) List all the genes you can find in the human genome that have a reasonable amount of sequence similarity with the Neanderthal sequence. Support your answer with data and cite your sources.

There were 3-4, depending where you cut off. Needed to use E-value to justify cut off (0.002 or thereabouts). It was a judgment call.

15 pts (5 pts each)

5) Find the three most similar non-human orthologs and support your answer with data and cite your sources. Tell me the common name for the 3 species you find. Align these three and the human sequences with the Neanderthal sequence. Show your alignment and then convince me that the reported Neanderthal sequence is not a modern contamination.

3 other primates, two chimp-like, one gorilla.

align them and then you would see that no modern source has the point mutation. Must be ancient.

8 pts

6) Final question on this paper.... Why does it make sense for you to get the results you did with question #5 above? What do you know, or you can easily find out, about the species you identified that validates your findings above?

For full credit, I was hoping you would comment on the hair color of the apes (dark) and the light hair color in the Neanderthal as indicated in the paper. Other MCRA answers got most of the credit.

>lcl|50570 Homo_sapiens_MC1R
Length=954

Score = 159 bits (86), Expect = 5e-44
Identities = 88/89 (98%), Gaps = 0/89 (0%)
Strand=Plus/Plus

```
Query 1   TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 60
          |||
Sbjct 847 TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 906
          |||

Query 61  CAGGAGCTCCGCGGGACGCTCAAGGAGGT 89
          |||
Sbjct 907 CAGGAGCTCCGCGGGACGCTCAAGGAGGT 935
          |||
```

>lcl|50571 Pan_paniscus_MC1R
Length=954

Score = 159 bits (86), Expect = 5e-44
Identities = 88/89 (98%), Gaps = 0/89 (0%)
Strand=Plus/Plus

```
Query 1   TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 60
          |||
Sbjct 847 TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 906
          |||

Query 61  CAGGAGCTCCGCGGGACGCTCAAGGAGGT 89
          |||
Sbjct 907 CAGGAGCTCCGCGGGACGCTCAAGGAGGT 935
          |||
```

>lcl|50572 Pan_troglodytes_MC1R
Length=954

Score = 159 bits (86), Expect = 5e-44
Identities = 88/89 (98%), Gaps = 0/89 (0%)
Strand=Plus/Plus

```
Query 1   TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 60
          |||
Sbjct 847 TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 906
          |||

Query 61  CAGGAGCTCCGCGGGACGCTCAAGGAGGT 89
          |||
Sbjct 907 CAGGAGCTCCGCGGGACGCTCAAGGAGGT 935
          |||
```

>lcl|50573 Gorilla_gorilla_MC1R
Length=951

Score = 159 bits (86), Expect = 5e-44
Identities = 88/89 (98%), Gaps = 0/89 (0%)
Strand=Plus/Plus

```
Query 1   TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 60
          |||
Sbjct 844 TTTCTCGCCCTCATCATCTGCAATGCCATCATCGACCCCTCATCTACGCCTTCCACAGC 903
          |||

Query 61  CAGGAGCTCCGCGGGACGCTCAAGGAGGT 89
          |||
Sbjct 904 CAGGAGCTCCGCGGGACGCTCAAGGAGGT 932
          |||
```

Read the Chivian *et al.* paper carefully and answer these questions. You may need to find some other web sources of information to fill in any gaps in your knowledge. Part of this test is to see whether you can access needed information effectively. I am supplying you with the supplemental materials, but do NOT print out this

8 pts (5/3 points)

7) Describe how the DNA was sequenced for this paper. Start after the DNA was isolated and conclude your answer with a detailed description of how they wound up with a single, complete genome of 2.35 Mbp from a single species. Describe the size and shape of this species and support your estimates with data.

Needed to address ddNTP and pyrosequencing methods.

Photo had size and scale bar. Range of size estimates between 2-4 μm .

8 pts

8) What surprising finding did they report relative to this organisms metabolism and interaction with its environment? What evidence supports their conclusion and what caveats to they offer with their interpretation?

N and C fixation, and SO₄ and reducing agent. Self-contained metabolism and ecosystem.

14 pts (4, 4, 4, 2 pts)

9) Produce a screenshot of a phylogenetic tree of all Archaea using one or more of the tools used in this paper. You do NOT need to install any software to perform this function. Using the same tool, provide a screenshot of a cluster of closely related species (bacterial or archaeal) who evolved recently and another screenshot of a group of species whose most MRCA was long ago. What species is missing that you had good reason to expect to find? Provide text to support all three of your screenshots.

Needed to use Microbesonline.org. This web site had all the information. The missing species was the one in the paper. It was not in the tree!!

5 pts (3/2 pts)

10) What is CRISPR and what role does it play in a prokaryote's life (do not simply repeat the phrase that appears in the same sentence that defines the acronym)? How can CRISPRs be used to help decipher genomes?

Immune system for microbes. Used to track evolution of species.

8 pts

11) What evidence did the authors supply to support their claim that the Sat gene has arrived in this genome by horizontal gene transfer? Do you believe them or not? Support your interpretation by citing evidence for your conclusion.

synteny, divergence of sequence between paralogs, and proximity to tRNAs as site of HGT.