

Feedback — Course Project **Please Note: No Grace Period**

[Help Center](#)

Thank you. Your submission for this exam was received.

You submitted this exam on **Mon 10 Aug 2015 9:10 PM PDT**. You got a score of **7.00** out of **10.00**. You can [attempt again](#), if you'd like.

In this project you will be reading a genomic data science paper and answering some questions to help you learn about how the different fields in genomic data science work together and to evaluate your understanding of some of the concepts we have learned throughout the course.

The paper we are reading is called: “Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss?” You can access the paper from the Science Magazine site here:

<http://www.sciencemag.org/content/292/5523/1903.full>.

You'll need to register for a free Science account to access the full text. If you aren't able to do that, then you can also access the via this link:

[Salzberg et al. 2001](#)

You may also find the annotated version of the manuscript helpful for the project.

[Annotated version](#)

Question 1

Why did the authors write this paper?

Your Answer

Score

Explanation

☐ To compute the E-values for the BlastP matches to the proteins from the human proteome.

☐ To propose the hypothesis that genes had been “laterally” transferred to humans.

☐ To identify species that share more genes in common between humans and bacteria.

☒ To propose a plausible alternative to the hypothesis that genes had been “laterally” transferred to humans. ✓ 1.00

Total	1.00 / 1.00
-------	----------------

Question 2

What is “lateral gene transfer”?

Your Answer	Score	Explanation
-------------	-------	-------------

☒ When genetic information is passed from bacteria to humans directly through infections or another process. ✓ 1.00

☐ When a gene is transferred out of the DNA and permanently lost.

☐ When genes are transferred out of the nucleus and into the cell.

☐ When genetic material is passed between one human and another human.

Total	1.00 / 1.00
-------	----------------

Question 3

Why is lateral gene transfer (LGT) from bacteria to humans unlikely?

Your Answer	Score	Explanation
-------------	-------	-------------

☒ There hasn't been sufficient time since humans and bacteria diverged for laterally transferred genes to spread through the population. ✖ 0.00

☐ Because human cells destroy bacteria too quickly for LGT to occur.

☐ Bacterial genes are so different from human genes that a human could not survive with bacterial genes in his/her chromosomes.

☐ Because a bacterium would have to infect a germline cell, enter the nucleus of that cell, and insert some of its DNA into one of the host's chromosomes, after which the mutation would then have to provide an evolutionary advantage to spread through the population.

Total	0.00 / 1.00
-------	----------------

Question 4

What are homologs?

Your Answer	Score	Explanation
-------------	-------	-------------

☐ Two genes in different organisms that have been mutated at the same rate.

☐ Two copies of a gene in different organisms that share a common ancestor.

☒ Genes that have identical DNA sequences ✖ 0.00

☐ Identical mutations that occurred over evolutionary time.

Total	0.00 / 1.00
-------	----------------

Question 5

What was the main method used to rule out lateral gene transfers between humans and bacteria?

Your Answer	Score	Explanation
<input type="radio"/> If a homolog of a eukaryote was found in another eukaryote but not in humans.		
<input checked="" type="radio"/> If a homolog of a gene found in humans was also found in a species of nonvertebrate eukaryotes.	✓ 1.00	
<input type="radio"/> If genes were found to have mutated between eukaryotic genomes and human genomes.		
<input type="radio"/> If a homolog of a gene found in humans was also found in bacteria.		
Total	1.00 / 1.00	

Question 6

Why would this method rule out lateral gene transfers?

Your Answer	Score	Explanation
<input type="radio"/> Inheritance of common genes is less common than lateral gene transfer and nonvertebrate eukaryotic organisms and humans are evolutionarily “closer” than bacteria and humans. If humans and bacteria share a homologous gene, it was likely directly passed from bacteria to humans.		
<input type="radio"/> Lateral gene transfer is a more likely explanation only when organisms that are close together evolutionarily share homologous genetic material.		
<input type="radio"/> Nonvertebrate eukaryotic organisms and bacteria are evolutionarily “closer” than invertebrate eukaryotic organisms and humans. If they share a homologous gene, then bacteria are likely to have passed genes directly to humans.		

- ☒ Genes that appear in nonvertebrate eukaryotic organisms and humans are genes that must have been shared in common well after humans and bacteria diverged, so these genes probably weren't transferred directly to humans. ✓ 1.00

Total	1.00 / 1.00
-------	----------------

Question 7

What are the biological, computational, and statistical parts of Figure 1?

Your Answer	Score	Explanation
-------------	-------	-------------

☐ **Biological:** the argument that a Blast cutoff of 10^{-10} should define homologs

Computational: The plotting of number of genomes versus number of genes in common

Statistical: Observing and quantifying the trend in genes shared versus genome sample size.

☐ **Biological:** the argument that a Blast cutoff of 10^{-10} should define homologs

Computational: The identification of homologs by performing Blastp searches on known protein sets.

Statistical: Observing and quantifying the trend in genes shared versus genome sample size.

☒ **Biological:** the argument that lateral transfer should be ruled out if there is a human/nonvertebrate eukaryote homologs. ✓ 1.00

Computational: The identification of homologs by performing Blastp searches on known protein sets.

Statistical: Observing and quantifying the trend in genes shared versus genome sample size.

☐ **Biological:** the argument that lateral transfer should be ruled out if there is a human/nonvertebrate eukaryote homologs.

Computational: the plotting of number of genomes versus number of genes in common

Statistical: Observing and quantifying the trend in genes shared

versus genome sample size.

Total	1.00 /
	1.00

Question 8

What are the biological, computational, and statistical parts of Figure 2?

Your Answer	Score	Explanation
-------------	-------	-------------

☐ **Biological:** the argument that proteins should have more similar sequences if they are evolutionarily closer.

Computational: The identification of homologs of human HAS genes by iterative BlastP searches and application of the neighbor-joining algorithm to create the phylogenetic tree.

Statistical: The inference that humans cluster more closely (have smaller distances to) other eukaryotes than to bacteria.

☐ **Biological:** the argument that proteins should have more similar sequences if they are evolutionarily closer.

Computational: The calculation of statistical significance of the protein hits in the Blastp search.

Statistical: The statistical modeling of protein sequences via a Markov Model.

☐ **Biological:** the argument that lateral gene transfer is less common than standard gene flow through reproduction

Computational: The storage of data in a low redundancy protein database.

Statistical: The inference that humans cluster more closely (have smaller distances to) other eukaryotes than to bacteria.

☒ **Biological:** the argument that lateral gene transfer is less common than standard gene flow through reproduction. ✖ 0.00

Computational: The calculation of statistical significance of the protein hits in the Blastp search.

Statistical: The statistical modeling of protein sequences via a Markov Model.

Total	0.00 /
	1.00

Question 9

The analysis in this paper required multiple data sources. Which of the following data sources was not used in the paper?

Your Answer	Score	Explanation
<input checked="" type="radio"/> The complete set of noncoding RNA genes from the human genome.	✓ 1.00	
<input type="radio"/> The complete set of genes from the fruit fly, nematode worm, yeast, and mustard weed genomes.		
<input type="radio"/> The complete set of annotated genes from two different versions of the human genome.		
<input type="radio"/> The set of all known genes (at the time) from the malaria parasite, <i>Plasmodium falciparum</i> .		
Total	1.00 / 1.00	

Question 10

In the end what is the conclusion of the paper?

Your Answer	Score	Explanation
<input checked="" type="radio"/> That a more plausible explanation for the observation of homologous genes found in bacteria and humans but not in non-vertebrate eukaryotes is gene loss and low sample size.	✓ 1.00	
<input type="radio"/> That the argument for lateral gene transfer is statistical because we must average over multiple possible transfer events.		
<input type="radio"/> That clustering organisms by protein similarity is the right way		

to discover lateral gene transfer events.

☐ That genes are more likely to be laterally transferred from certain types of bacteria to humans.

Total

1.00 /

1.00